ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business

When sustainability matters: performance dynamics of European ESG ETFs

Author:	Simone Patarchi
Student number:	699233
Thesis supervisor:	Dr. Jan Lemmen
Second reader:	Dr. Kan Ji
Finish date:	July 2024

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics, or Erasmus University Rotterdam.

Abstract

Environmental, Social, and Governance (ESG) integration in the Exchange-Traded Funds (ETFs) market has obtained significant traction in recent years. This study aims to evaluate the risk-return profile of ETFs, to understand whether ESG factors improve their performance. Several ESG scores are employed in regressions to understand their relation with four performance measures. Moreover, portfolios are created from the Morningstar Sustainability Ratings and employed in multifactor models to detect any sign of underpricing. Some overperformance is found in relation to the most sustainable portfolios, with a positive significant correlation between sustainability (especially the environmental pillar) and the performance measures. Geographically, overperformance is consistent only in the North American region, with some evidence also on the Global stage, probably influenced by the former.

Keywords: ESG ETFs, Risk-adjusted performance, Sustainability, European-traded ETFs, Multifactor models

JEL Classification: G11, G12, G15, Q01

Contents

List of Tables

List of Figures

1 Introduction

With the world experiencing a constant evolution, financial markets have always adapted to better suit investors' needs. During the last few decades, our society experienced a massive technological and ethical advancement that resulted in a more globalized and sustainable world. These improvements also heavily impacted the state-of-the-art technologies used in financial markets, leading to new and innovative methods to invest in them in an ever easier and faster fashion. ETFs and ESG rankings can be considered two of the most notable recent developments, and this thesis aims to study their combined behavior with a simple research question in mind: *"Does the integration of environmental, social, and governance (ESG) criteria into European ETFs enhance financial performance relative to traditional ETFs?"*.

While tons of literature regarding ESG funds and stocks can be found, only a limited number of studies handle the topic of ESG ETFs, discovering conflicting findings. Most of these studies also included the early adopters of ESG criteria, which, given their market concentration and rudimentary ESG implementation strategies, could most probably show anomalous behaviors compared to the more recent (and advanced) ESG ETFs. Another widely adopted issue is the method of comparison to evaluate possible underpricing, as many studies compare sustainable ETF returns with benchmarks, completely disregarding the fact that overperformance may not be a direct consequence of the ETF's sustainability profile.

My contribution to the extant literature is primarily the study of ESG ETFs in all the European markets (which was never done before, with research mainly concentrating on the US or specific European countries such as Rompotis, 2023). Moreover, to tackle the two important issues listed before, I have constructed a sample consisting of the last five years of daily returns to study the current and more developed ETF ESG market, including two important downturns (namely the COVID-19 crisis and the 2022 stock market decline) to verify its robustness, and compared using risk-adjusted metrics the ETFs deemed sustainable with their counterparts showing low ESG ratings (and not with general benchmarks) to exclude from the analysis any overperformance that other ETF characteristics may cause.

The results of this research highlighted a few interesting remarks, such as a significant negative relation between performance measures and the magnitude of ESG scores, but there are contrasting results on its robustness during downturns. When splitting the ESG score into its singular components, the pillar that mainly drives the returns is the Environmental one, followed by the Governance pillar that however shows an inverse relation with sustainability. When splitting the ETFs using their *Morningstar Sustainability Rating*, the *High* bracket shows some abnormal returns, accounting for approximately between 5.7% and 7.3% yearly. By analyzing this overperformance geographically, North America is found to be the only macroregion (along with the Global one) to showcase significance in this regard.

The thesis is structured as follows: Section 2 delineates the theory and previous studies on the topic, Section 3 introduces the type of data employed and their statistical summaries, Section 4 explains the methodology followed during the analysis, Section 5 displays the main findings, and Section 6 offers some insights and concludes.

2 Theoretical Background

This section is devoted to the introduction of the topics that will be discussed in the subsequent analysis. In the first place, some background regarding ETFs and ESG ratings is given to the reader to get him accustomed to these tools, setting the stage for a more technical literature review of the previous studies concerning this matter. Based on the review, a series of hypotheses are formulated to guide the analysis of the research question from multiple different perspectives.

2.1 Exchange-Traded Funds

Exchange-Traded Funds (usually referred to just as ETFs) are a type of investment fund with the peculiarity of being traded on stock exchanges. Even though early experiments of traded funds can be traced back to 1989 with the *Index Participation Shares (IPS)*, the first ETF was created in 1993 with the name of *Standard & Poor's Depositary Receipts* (now *SPDR S&P 500 Trust ETF* - SPY), developed to track the *S&P 500 Index Fund*. In this early stage, ETFs held very little significance and remained largely overshadowed by the general investor, at least until the sharp growth in popularity that started after the Global Financial Crisis, making the ETFs market experience an enormous boost that brought it from 0.9 trillion dollars in 2009 to over 10 trillion in 2024 (ETFGI, 2024). Consequently to the expansion, the market kept growing and creating new types of products tailored to accommodate every type of customer, ranging from actively managed to index-replicating ETFs comprising most asset classes such as stocks, bonds, futures, debts, currencies, and commodities. As of now, ETFs are a quick and effective way to take advantage of the diversification benefits of a fund without giving up the accessibility and liquidity that an ordinary fund cannot offer.

2.2 ESG ratings

In recent years, sustainability has acquired more and more relevance in the political and economic landscape, consequently becoming a crucial subject also in investing practices. In order to assess these issues at a company level, financial practitioners developed a score that serves to determine the firm's position in relation to several issues concerning sustainable, ethical, and social problems. There is no general consensus on a single ESG score, as every rating agency uses slightly different criteria and models for the assessment. Nevertheless, a few rating firms (such as *MSCI ESG*, *Sustainalytics*, and *S&P ESG*) imposed themselves over the competitors as more reliable in depicting the real state of the company. The score is mainly divided into three different pillars, each one representing key issues relevant to the investor:

2.2.1 Environmental (E)

The Environmental pillar puts a major emphasis on climate change, one of the issues that will surely affect living and operating in the near future, as extreme events given by the abrupt increase of environmental pollution (in the form of gas emissions, usage of non-recyclable materials, etc.) and biodiversity risks (human-driven extinctions, anthropogenic exploitation of biomes, etc.) already started showing negative externalities that heavily affect companies' profitability by reducing revenues and introducing new climate-related costs.

2.2.2 Social (S)

The second pillar, Social, is focused on handling issues related to ethical principles and rights, as these are also widely considered by investors. These behaviors may lead to inefficient productivity as well as legislative problems given the shady legislative area in which most of these are classified. The major drivers of this pillar are lack of employee welfare (unsafe working conditions, no diversity, disparity of treatments, etc.), geopolitical issues (conflicts, humanitarian crises, violation of human rights, etc.), absent consumer protection and animal exploitation (animal testing, forced breeding, intensive and unethical exploitation).

2.2.3 Governance (G)

Governance plays a big role in determining the success or failure of a company, as it serves to ensure companies to be more accountable, resilient, and transparent to investors, ultimately giving them the tools to respond to stakeholder concerns. It mainly handles the firm's organizational structure and processes, ranging from executives (compensation, distribution of powers, behaviors, and diversity) to employees (fair wage, workplace relations, possible discriminations), and ethical issues (correct business practices, transparency and prevention of bribery and corruption).

2.2.4 Morningstar Sustainalytics

In this thesis, the emphasis will be given to the ESG Rating issued by Sustainalytics, a Morningstar company that offers high-quality ESG data and occupies a leading position in the ESG rating market. In a joint effort with Morningstar, the company in 2016 released the first *Morningstar Sustainability Rating*, ranking mutual funds and ETFs on the basis of their ESG profile. This rating is peculiar as it does not rank the ESG on a positive scale where a higher value is better, instead, it focuses on the negative, "risky" contribution by quantifying an ESG Risk

Score, where high numbers are symptoms of higher risks related to the three pillars. Below, in Table 1, it can be found the weights that each pillar has in the general score contribution.

Category	Issue	Contribution
Environmental (43.3%)	Carbon - Own Operations	19.2%
	Resource Use	10.3%
	Emissions, Effluents and Waste	7.1%
	Environmental and Social Impact of Products and Services	6.7%
Social (34.1%)	Human Rights	22.8%
	Occupational Health and Safety	7.5%
	Community Relations	3.8%
Governance (22.6%)	Corporate Governance	11.9%
	Business Ethics	6.7%
	Human Capital	4.0%

Table 1: Sustainalytics ESG risk score contribution

2.3 Literature Review

While studies on ESG overperformance over the last couple of years have become a very common topic throughout the different asset classes, ESG ETFs are quite a recent phenomenon, hence only a handful of research cover this topic.

Pavlova and de Boyrie (2022) aims to assess whether COVID-19 affected ESG ETFs' profitability, finding out that lower-rated ESG ETFs outperformed higher-rated ones and the market in the pre-COVID period. However, during COVID-19 sustainable ETFs showcased non-significant negative alphas indicating the lack of a stronger downturn protection by higher-rated ETFs, but still yielding a better result than the market. Other studies, such as Folger-Laronde et al. (2020) have also investigated the same period, with results similar to the former paper, finding that ESG characteristics of an ETF do not shield from losses during periods of crisis, with financial performance not directly related to sustainability performance. Instead, Nguyen (2023) found results contradicting the precedent ones, with an increased performance of sustainable ETFs compared to the market, depicting these instruments as more resilient to shocks. Moreover, the study revealed ETFs' excellent index-tracking capabilities when subjected to systematic shocks. Other studies covering broader periods showcase mixed results: Meziani (2020), covering ESG ETFs with more than 10 years of market presence, found modest risk-adjusted return underachieving compared to the benchmark even though competitively priced (given the lack of premium paid by the investor). Kanuri (2020), by employing portfolios of equally-ranked ETFs and comparing them to two benchmarks (the *Russell 3000 ETF* - IWV and the *SPDR Global Dow ETF* - DGT), discovered a general ESG underperformance though not constant over time, as during the Global Financial Crisis both ESG portfolios did better in risk-adjusted terms by yielding lower risk-adjusted losses (but higher absolute cumulative losses given the much higher volatility). Focusing on passive ESG ETFs, Dumitrescu et al. (2022) found some underperformance compared to the S&P 500, however showing that by only considering positive screening (rather than negative screening) outperformance of the benchmark is found. The same research also found that increased industry competition can be beneficial for increasing performance. Rompotis (2022) examines 49 UK ETFs deemed sustainable comparing them against their own benchmark and the *FTSE 100*, eventually discovering a slight underperformance for the former while overperformance to the latter. Moreover, the study found a negative relationship between ESG metrics and financial performance, hinting that responsible investing entails lower returns. A further study by the same author (Rompotis, 2023) investigates ESG ETFs in the US using the *S&P 500* as the benchmark, concluding that sustainable ETFs outperform the market index both in absolute terms and risk-adjusted ones, however on average these ETFs yield a higher risk than the benchmark. After having analyzed ESG ETFs, also some considerations on the general market of Exchange-Traded Funds are important: Bhattacharya et al. (2016) discovered that the individual investor is not subjected to any benefit by investing in ETFs, mainly due to wrong timing and ETF selection, being biased towards excessively specialized ETF with much higher fees than low-cost well-diversified alternatives. Thus, it is of crucial importance not only to study sustainable ETFs' behavior but also the ones not complying with ESG standards must be taken into consideration, to exclude every possible effect given by the ETF itself.

Below, in Table 2, can be found a more extensive and detailed list of the relevant papers, also including further research on sustainable assets risk/return profiles. Apart from a few outliers, most of the literature shares common traits in sketching the methodology: given the crucial importance of adjusting for the riskiness of the asset, risk-adjusted ratios such as the ones developed by Sharpe, Sortino, Traynor, and other measures accounting for a risk correction such as M^2 (otherwise known as "Modigliani risk-adjusted performance" or "RAP") and Jensen's alpha are the main strategies employed. Regarding the latter, several different single- and multifactor models are used, starting from simpler ones (e.g. the CAPM by Sharpe, 1964) to more complex frameworks likewise the model developed by Fama and French (2015).

2.4 Hypotheses

The analysis spans the widely debated topic of sustainable assets' performance by specifically establishing whether sustainable ETFs yield significantly higher returns than their non-sustainable counterparts after correcting for risk, hence a research question can be formed: *"Does the integration of environmental, social, and governance (ESG) criteria into European ETFs enhance financial performance relative to traditional ETFs?"*.

Under this broader research objective, in order to study from multiple different perspectives the risk-return profile of ESG ETFs and considering the interesting insights of previous research, several smaller research hypotheses are made:

1. *ETF profitability is not significantly correlated with the magnitude of sustainability scores.*
2. *Governance drives the ETF performance more than Environmental and Social scores.*
3. *There is no significant overperformance of ETFs falling into higher sustainability brackets.*
4. *The ESG ETF performance is not significantly different among investment regions.*

Author(s) (Year)	Time period	Region	Asset class	Methods		ESG effect?	Other results
Derwall et al. (2011)	Jan'92 - Dec'98	US	Employee-relations weighted stock portf	Jensen's α	Car	Positive (SR) Non sign.(LR)	Shunned-stock hypothesis holds
Nofsinger and Varma (2014)	Jan'00 - Dec'11	US	ESG funds and ETFs	Jensen's α	CAPM, F3, Car	Non sign.	Better positive than negative screening
Kanuri and McLeod (2016)	Jun'02 - May'14	US	EW and VW "wide moat" stock portf	Ratios Jensen's α	Sh, So, Tr, Ω UP, M^2,$M^2\alpha$ CAPM, Car, 5F	Positive	
Polbennikov et al. (2016)	Jan'07 - Dec'15	Global	Active funds	Regression and tracking error		Positive	G main driver
Dolvin et al. (2017)	Jan'12 - Dec'16	US	EW and VW funds portf	Jensen's α	Car	Non sign.	High scores usually given to large cap
Filbeck et al. (2019)	Aug'09 - Dec'16	US, Canada	VW ESG stock portf	Ratios Jensen's α	Sh, Tr CAPM, 3F, Car, 5F	Positive	G positively related E negatively related
Folger-Laronde et al. (2020)	Jan'20 - Mar'20	Global	ESG ETFs	ANOVA Turkey test		Non sign.	SRI not resilient to crisis
Kanuri (2020)	Feb'05 - Jul'19	US	EW and VW ETF portf	Ratios Jensen's α	Sh, So CAPM, Car, 5F	Negative	Outperformance in bull markets
Meziani (2020)	Jul'09 - Jul'19	US	ESG ETFs	Ratios Jensen's α	Sh, Tr, So, IR CAPM	Non sign.	
Pástor and Vorsatz (2020)	Jan'17 - Apr'20	US	Active funds	Jensen's α	CAPM, F3, Car, F5, F5 + Mom	Positive	Active management underperforms
Kanamura (2021)	Aug'09 - Dec'16	US, Canada	HY bond ETFs	Classic to ESG ETF trading model		Positive	
Dumitrescu et al. (2022)	Jan'10 - Dec'20	US	EW ESG ETF portf	Jensen's α	CAPM, F3, Car, F5	Non sign.	
Pavlova and de Boyrie (2022)	Nov'19 - May'20	Global	ETFs	Jensen's α	CAPM, F3, Car, F5, F5 + Mom	Non sign.	
Rompotis (2022)	Jul'07 - Dec'20	UK	ESG ETFs	Ratios Jensen's α Regression	Sh, Tr CAPM, F5	Mixed	
Nguyen (2023)	Jan'19 - Mar'22	US	EW ETF portf	Jensen's α	CAPM, F3, Car, F5, F5 + Mom	Positive	ETFs good for tracking indices
Rompotis (2023)	Jan'19 - Dec'21	US	ESG ETFs	Ratios Jensen's α Regression	Sh, Tr, So, IR CAPM, F5 + Mom	Non sign.	

The table provides an overview of the main research papers relative to the studied phenomena. Due to space constraints, the following abbreviations are used throughout the table: **EW** = equally-weighted, **VW** = value-weighted, **HY** = high yield, **Sh** = Sharpe ratio, **So** = Sortino ratio, **Tr** = Traynor ratio, Ω = Omega ratio, **UP** = Upside Potential ratio, $\mathbf{M^2}$ = Modigliani risk-adjusted performance, $\mathbf{M^2}\alpha$ = Modigliani risk-adjusted performance alpha, **CAPM** = Capital Asset Pricing Model, **3F** = Fama and French (1992) model, **Car** = Carhart (1997) model, **F5** = Fama and French (2015) model, **F5 + Mom** = Fama and French (2015) plus momentum model, **SR** = short run, **LR** = long run

Table 2: Meta table

3 Data

To verify the aforementioned hypotheses, data regarding returns, ESG scores, and fund characteristics of the relevant ETFs is needed. Morningstar collects and provides high-quality ETF information, hence the usage of its proprietary database, Morningstar Direct, is a logical choice. Given the focus of the analysis being ETFs traded on European markets during the last five years, the collection of data is performed through the function *Advanced Research*. The filters used in this screening procedure to exclude unwanted ETFs are:

- *Exchange Country: all European countries*
- *Inception Date: before 30/04/2019*
- *Fund Size: over 500 million USD*
- *Morningstar Sustainability Rating: exclude non-rated ETFs*

Now, having the list of ETFs that satisfy all criteria, the following data points are collected:

- *Returns (from 01/05/2018 to 30/04/2024)*
- *Fund Size (on 30/04/2024)*
- *Annual Net Expense Ratio (on 30/04/2024)*
- *Investment Area (on 30/04/2024)*
- *Morningstar Sustainability Index (on 04/2024)*
- *Historical Corporate Sustainability Score (on 04/2024)*
- *Portfolio Corporate Sustainability Score (on 04/2023, 10/2023, 04/2024)*
- *Portfolio Environmental Risk Score (on 04/2023, 10/2023, 04/2024)*
- *Portfolio Social Risk Score (on 04/2023, 10/2023, 04/2024)*
- *Portfolio Governance Risk Score (on 04/2023, 10/2023, 04/2024)*
- *Portfolio Unallocated Risk Score (on 04/2023, 10/2023, 04/2024)*

Morningstar Sust. Rating	**Mean**	**St.Dev.**	**Median**	**Min.**	**Max.**	**Range**	**Skew.**	**Kurt.**	**Fund Size**	**Net Exp. Ratio**
High	0.05	0.65	0.05	-1.30	1.40	2.71	0.01	0.47	2.29	0.28
Above Average	0.04	0.74	0.04	-2.00	2.16	4.16	0.08	3.21	2.17	0.30
Average	0.04	0.62	0.04	-1.91	2.07	3.98	0.08	3.98	4.77	0.30
Below Average	0.04	0.63	0.03	-1.55	1.71	3.26	0.05	1.57	3.25	0.28
Low	0.03	0.72	0.03	-1.06	1.15	2.21	0.03	-0.80	1.55	0.40

This table showcases the daily returns statistics for each Morningstar Sustainability Rating. The values are computed as the average of the ETFs falling in that category. Returns and Net Expense Ratios are in percentage terms. Fund Sizes are displayed in Billion USD.

Table 3: Daily returns and fund characteristics statistics

The aforementioned variables were collected for the 280 ETFs matching the desired characteristics. No NA values were generated apart from a few missing Annual Net Expense Ratios. To address this issue, the average value of the remaining is used wherever no value is found.

The daily returns statistics can be found in Table 3, split in the five different *Morningstar Sustainability Ratings*. The mean return is slightly higher for superior ratings, however, it can also be seen that the biggest standard deviation is achieved within the *Above Average* bracket, also subject to the highest range. Meanwhile, as expected, the largest and more generic ETFs have an *Average* rating, with funds falling in the other brackets probably being more specialized, thus attracting lower investments. It is also worth mentioning that the average Net Expense Ratio for the *Low* category is much higher than that of the other brackets.

Regarding the ESG metrics, Table 4 contains the entirety of variables used in the several steps of the analyses. For some also the time series of the data point was available for further investigations using present and past scores. On average, the highest scores (i.e. worse ESG performances) are registered by the 6-month lagged observations, followed by the 1-year lag, indicating a general increase in scores during the last 6 months, as opposed to what happened in the semester before.

On top of the previously mentioned variables, also the time series of the factors of interest are needed to perform multifactor analysis. These can be easily downloaded from Kenneth French's website (French, 2024), where daily factor loadings are available.

Variable	Obs.	Mean	St.Dev.	Median	Min.	Max.	Range	Skew.	Kurt.
Historical Corporate Sustainability Score	280	21.02	3.18	20.98	11.00	34.79	23.79	0.82	3.96
Portfolio Corporate Sustainability Score	280	20.85	3.19	20.77	10.69	34.60	23.91	0.84	4.08
Portfolio Corporate Sustainability Score (1y lag)	280	21.13	3.22	21.04	11.52	35.11	23.59	0.91	3.77
Portfolio Corporate Sustainability Score (6m lag)	280	21.08	3.16	21.05	10.97	34.72	23.75	0.75	3.91
Portfolio Environmental Risk Score	280	4.70	2.41	4.34	0.27	18.11	17.84	3.03	13.19
Portfolio Social Risk Score	280	8.35	1.70	8.68	0.28	11.87	11.59	-1.83	5.63
Portfolio Governance Risk Score	280	6.78	1.23	6.86	0.20	11.39	11.19	-1.36	7.62
Portfolio Unallocated Risk Score	280	1.02	3.08	0.03	0.00	24.16	24.16	5.02	29.35
Portfolio Environmental Risk Score (1y lag)	280	4.73	2.39	4.41	0.27	18.16	17.89	2.94	12.60
Portfolio Social Risk Score (1y lag)	280	8.40	1.68	8.66	0.44	12.03	11.59	-1.86	6.07
Portfolio Governance Risk Score (1y lag)	280	6.94	1.28	7.04	0.27	11.26	10.99	-1.37	7.33
Portfolio Unallocated Risk Score (1y lag)	280	1.06	3.19	0.05	0.00	24.58	24.58	5.04	29.46
Portfolio Environmental Risk Score (6m lag)	280	4.77	2.41	4.40	0.28	18.19	17.91	2.99	13.12
Portfolio Social Risk Score (6m lag)	280	8.42	1.71	8.73	0.46	11.93	11.47	-1.84	5.56
Portfolio Governance Risk Score (6m lag)	280	6.87	1.25	6.86	0.31	11.24	10.93	-1.38	7.28
Portfolio Unallocated Risk Score (6m lag)	280	1.02	3.12	0.00	0.00	24.14	24.14	5.00	28.97

Presented above are the relevant statistics for the sustainability measures employed in the analysis. The variables without a date represent the "current" value (i.e. the value observed during the last month of the sample). If the variable is referring to a previous period, the date can be found between brackets.

Table 4: ESG summary statistics

4 Method

Apart from the data gathered in the previous step, some other metrics are needed to explore the first two hypotheses. First of all, cumulative returns must be computed to investigate the cross-section of the available ETFs. The 6-month and 1-year returns are chosen to represent longer periods, even though it is crucial to highlight that these thresholds are completely arbitrary. Starting with daily returns, the following formula is used:

$$R_T = \left(\prod_{t=x}^{n} (1 + R_t) \right) - 1 \quad \text{with} \ \ x = \{n-129, n-259\} \tag{1}$$

The two different x specifications allow us to compute the two requested intervals, the 6-month ($n-129$, since there are 130 weekdays in a semester) and the 1-year ($n-259$) ones. Moreover, Sharpe ratios are essential to understand better the risk-adjusted returns and can be easily computed using Equation 2:

$$\text{Sharpe ratio} = \frac{R_T - R_f}{\sigma_T} \tag{2}$$

With R_T being the return of the ETF, σ_T the corresponding volatility and R_f the risk-free rate in that specific period. The last important step is testing whether these metrics are significantly different from zero, hence relevant for statistical purposes. This is done through the Jobson-Korkie test (the corrected version by Memmel, 2003), which given its simplicity, computational efficiency, and low number of assumptions in the last two decades has become the standard practice when dealing with Sharpe ratios for scholars and practitioners. Here, the test statistic is modified to accommodate this type of analysis:

$$\begin{gathered} Z_{JK} = \frac{S\hat{R} - 0}{\sqrt{\hat{V}}} \\ \hat{V} = \frac{1}{T}(2 + \frac{1}{2} S\hat{R}^2) \end{gathered} \tag{3}$$

This is slightly different from the original Jobson-Korkie test statistic, as here is not used to compare two Sharpe Ratios (as it was originally intended), but is employed to assess the significance of a Sharpe Ratio being different from zero. After having collected all the necessary variables, it is essential to develop a robust methodology in order to guarantee theoretical accuracy and reliability. The previously mentioned hypotheses (Subsection 2.4) can be used as a starting point to subdivide the analysis.

4.1 Hypothesis 1: Sustainability impact

ETF profitability is not significantly correlated with the magnitude of sustainability scores.

In light of the fact that the first hypothesis focuses on the value and not on the final classification, this subsection uses *scores* (i.e. a quantitative evaluation derived from computations) and not *ratings* (i.e. a categorical or ordinal assessment grouping different scores in two or more categories). According to Sustainalytics ESG rating of corporate entities, two main scores contribute to the rating procedure: the *Portfolio Corporate Sustainability Score* (PCSS) and the *Historical Corporate Sustainability Score* (HCSS).

The first one, the PCSS, is defined by Morningstar as *"an asset-weighted average of Sustainalytics' company-level ESG Risk Scores"*. This, as all the other ESG Risk measures can take values in the range of 0-100, where a lower score indicates safer companies less exposed to these risks. Using this metric, we can construct the first cross-sectional OLS model:

$$\text{Perf}_i = \gamma_1 + \gamma_2 \text{PCSS}_i + \varepsilon \tag{4}$$

It is also necessary to note that multiple performance indicators are considered as dependent variables (*Perf*), namely 6-month cumulative returns, 1-year cumulative returns, 6-month Sharpe ratios, and 1-year Sharpe ratios, hence four different regressions are needed for each summarized version containing *Perf*.

The main drawback of this first regression is that the PCSS employed is at time n (i.e. last day of the observed period), and even if significant changes in scores usually take very long periods, some smaller adjustments may still happen, also influencing our returns of interest. Given this problem, the usage of a lagged PCSS corresponding to the starting date (that depending on the performance measure can be 6 or 12 months prior) can be used both as a new investigation and as robustness for the previous regression in case of similar results. The next model is consequently built as showcased below:

$$\begin{gathered} \text{Perf}_i = \gamma_1 + \gamma_2 \text{PCSS}_{i,x} + \varepsilon \\ \text{with} \;\; x = \{n-129, n-259\} \end{gathered} \tag{5}$$

As stressed also before, the notation *n-129* and *n-259* refer to the possible lags, as the two values represent the number of weekdays in six months and one year.

The second key metric, the HCSS, is an evolution of the first two variables used, summarizing the sustainability profile of the ETF during the last year. This is computed by means of a weighted average of the scores pertaining to the trailing 12 months, weighted on its proximity to the newest observation, consequently emphasizing newer observations. In the next regression, this measure is employed to understand its relation to the performance factors:

$$\text{Perf}_i = \gamma_1 + \gamma_2 \text{HCSS}_i + \varepsilon \tag{6}$$

After having considered these two variables separately, the next logical step is to mix them together in a single, comprehensive regression:

$$\text{Perf}_i = \gamma_1 + \gamma_2 \text{HCSS}_i + \gamma_3 \text{PCSS}_i + \varepsilon \tag{7}$$

After having established these techniques, is now time to make some general considerations on the important coefficients and tests employed. In these regressions, the emphasis is devoted to the coefficients of the regressors (γ_2 and γ_3) that express the relationship between sustainability and performance measures by studying their sign and magnitude. In addition to that, all the coefficients are tested for being significantly different from zero using two-tailed t-tests and verifying whether the associated p-values fall in the significance region. However, it is fundamental to stress that finding the determinants of the four performance indicators is not the aim of the research, as that would be a completely different (and very extensive) study. Having said that, omitted variables may have a role in altering the coefficient by enhancing its magnitude if correlated with our regressors, but since our focus is on the type of effect (i.e. the sign of the coefficient), indicating a positive or negative correlation, this does not invalidate our analysis.

4.2 Hypothesis 2: E, S, and G impact

Governance drives the ETF performance more than Environmental and Social scores.

This hypothesis further explores the determinants by separating the PCSS into its three pillars: Environmental, Social, and Governance. Similarly to before, the analysis employs the scores relative to these three metrics, namely the *Portfolio Environmental Risk Score* (E), the *Portfolio Social Risk Score* (S), and the *Portfolio Governance Risk Score* (G). Apart from these, there is also a fourth risk measure, the *Portfolio Unallocated Risk Score* (U), that comprises the aggregated score relative to the holdings for which there is no clear split into the three specific categories. As for the previous metrics, also these can take values from 0 to 100, where lower is better, though most scores range from 0-25. Using the four measures as dependent variables, the regression is thus structured as follows:

$$\text{Perf}_i = \gamma_1 + \gamma_2 \text{E}_i + \gamma_3 \text{S}_i + \gamma_4 \text{G}_i + \gamma_5 \text{U}_i + \varepsilon \tag{8}$$

Even in this case, the same line of reasoning used before is still valid: while the latest value can still be a good proxy, further investigation using the lagged values can bring new information and robustness to the discussion. The next regression therefore incorporates the lagged metrics:

$$\begin{gathered} \text{Perf}_i = \gamma_1 + \gamma_2 \text{E}_{i,x} + \gamma_3 \text{S}_{i,x} + \gamma_4 \text{G}_{i,x} + \gamma_5 \text{U}_{i,x} + \varepsilon \\ \text{with} \;\; x = \{n-129, n-259\} \end{gathered} \tag{9}$$

Depending on the performance measure, the corresponding lag must be used, as highlighted by the two possible values that x can take. Lastly, the coefficients and tests employed before are still valid to be used here: the coefficients (γ) give the type of the relation and the importance of the regressor, while the two-tailed t-tests yield the significance of the results.

4.3 Hypothesis 3: Rating-specific performance

There is no significant overperformance of ETFs falling into higher sustainability brackets.

The third hypothesis diverts attention from correlations to risk-adjusted performances, delving into the most critical part of the study. To verify this claim, a criterion to split our sample is needed. The most efficient method, both in terms of simplicity and effectiveness, is to use the *Morningstar Sustainability Ranking*. This raking system is built using the measures utilized in Section 4.1 together with the peer group sustainability scores. The distribution of the scores in these peer groups is then normalized to divide them into five different ESG categories: *High* (best 10%), *Above average* (next 22.5%), *Average* (next 35%), *Below average* (next 22.5%), and *Low* (worst 10%).

Having identified the criterion, this can now be used to split the sample into five distinct portfolios, each corresponding to a different rating. After checking that every portfolio would contain at least an acceptable number of ETFs, it is possible to start the splitting process. Both equally-weighted and value-weighted portfolios are created to assess potential similarities and differences in results. Given the two weighting methods, for the former a normal cross-sectional average of the daily returns for every ETF in that portfolio is used, while for the latter is needed a cross-sectional average weighted on size, making bigger ETFs more influential than their smaller counterparts. After having computed all the daily returns for every portfolio, multifactor models can be employed to understand whether there are instances of abnormal returns (i.e. underpricing caused by anomalies not priced by the systematic risk factors) or particular loadings on various anomalies. These models found wide applicability in the outstanding literature (see Section 2.3), becoming one of the most efficient ways to detect abnormal performance in every asset class. There are a number of different possible specifications to choose from, and in this analysis, the following will be used:

- **Sharpe (1964) CAPM model**

  The oldest factor model, which contains a single factor, the market premium:

$$R_i - R_f = \alpha + \beta(R_M - R_f) + \varepsilon \tag{10}$$

  The equation is composed of three main blocks: $R_i - R_f$, the portfolio excess returns, α, the aforementioned abnormal return, and $\beta(R_M - R_f)$, the market premium and its associated coefficient, beta, indicating the sensitivity to price changes respective to the market. This is the simplest factor model, however, due to its straightforwardness, it is still widely used even though the omitted variables may alter the β coefficient, shifting it from the real value.

- **Carhart (1997) four-factor model**

  Carhart first introduced this version by enhancing the previous Fama and French (1992) model with a momentum factor:

$$R_i - R_f = \alpha + \beta_M(R_M - R_f) + \beta_{SMB}\text{SMB} + \beta_{HML}\text{HML} + \beta_{UMD}\text{UMD} + \varepsilon \tag{11}$$

  Apart from the market factor used also in CAPM, three anomalies are used to price the asset: the first two, size (SMB) and value (HML), were also part of the Fama and French (1992) model. The size anomaly refers to the fact that smaller companies tend to outperform larger ones on a risk-adjusted basis, while value refers to the phenomenon that value stocks (those with low price-to-book ratios) tend to outperform growth stocks (those with high price-to-book ratios) over the long term. The last one, the momentum factor (UMD), is instead related to the assumption that past winners will continue to register positive returns in the near future and vice versa, and was a later addition by Carhart, aiming to improve even further the pricing model by implementing a third, very influential, anomaly.

- **Fama and French (2015) five-factor model**

  Also in this case the model is built starting from the Fama and French (1992) model, further introducing fundamental anomalies that may not be directly priced in by the market. The regression of this multifactor model is therefore specified as:

$$R_i - R_f = \alpha + \beta_M(R_M - R_f) + \beta_{SMB}\text{SMB} + \beta_{HML}\text{HML} + \beta_{RMW}\text{RMW} + \beta_{CMA}\text{CMA} + \varepsilon \quad (12)$$

  Compared to the previous two, this model incorporates the profitability (RMW) and investment (CMA) factors. Both anomalies stem from accounting measures, as the first refers to the observation that more profitable firms register higher stock returns, while the latter deals with the tendency for companies that have a conservative investment style to outperform those that invest aggressively. Even though the Efficient Market Hypothesis (Fama, 1970) would assume that all these characteristics should be already priced in, numerous studies show the existence of premia associated with these anomalies.

To use these three multifactor models, a series of variables are needed, starting from risk-free rates, market premia, and all the necessary factors. On Kenneth French's website (French, 2024), these are freely downloadable in daily, monthly, and yearly formats, alongside region-specific ones. The next logical step is to take every portfolio (from the most to the least sustainable ones) and use their daily returns to assess factor loadings and abnormal returns. To establish significance, the coefficients are then employed to carry out two-sided t-tests, eventually finding which results outline a coherent behavior and which are not consistent enough to be significant. Among all the coefficients, however, the most interesting one is α, depicting whether there is any sign of over- or underpricing, even though by hypothesis no significance related to it is expected.

4.4 Hypothesis 4: Region-specific performance

The ESG ETF performance is not significantly different among investment regions.

The last hypothesis deals with geographical differences: the goal is to understand whether the behavior detected in Chapter 4.3 is consistent throughout different areas of investment or if there are regions more or less susceptible to sustainability factors. Likewise to the previous hypothesis, also this one entails the division of the sample into smaller subsamples, but in this case, the splitting criterion is different: while before the division was based on ESG, now the characteristic to be studied is the location of the investment. The first problem arises in the regional classification: since every ETF has its purposes and interests, the more specialized ones will have peculiar regions of investments, therefore having no peers to construct a portfolio with. To counter this excessive fragmentation, I grouped these into broader regions, in order to have at least 10 ETFs falling in a single portfolio. This process resulted in five broad regions: Europe (the one with the most observations, accounting for almost 40% of the total), North America, Asia, Emerging (comprising all the ETFs invested in global emerging countries or only in specific emerging regions), and Global (representing only ETFs with investments in several different developed countries spread across the globe). Similarly to before, equally-weighted and value-weighted portfolios for the top ESG scores in every broad region are then created for further investigation and robustness of results, and the same multifactor models of the previous section (i.e. Equations 10, 11, and 12) can be still used to assess whether these portfolios show any sign of abnormal returns. Given the different regions, the same factors cannot be used and must be substituted by region-specific factors. Fortunately, also these are available on Kenneth French's website (French, 2024). To verify significance two-sided t-tests will be used.

4.5 Robustness checks

To ensure the reliability of the results, robustness checks are carried out by implementing the costs of investing on each ETF in the sample. Even though these costs are formed by several different entries, every ETF is required to disclose in the prospectus its *Gross Expense Ratio*, defined as the percentage of the assets that are expected to be paid in the next 12 months for interest and dividend expenses, operating expenses, and management fees. From this measure Morningstar derives the *Annual Report Net Expense Ratio* by adding back any fee waivers and reimbursements. Since these costs are paid on a daily basis, this metric can be employed to implement fund fees in the returns by first dividing it by 260 (i.e. the weekdays in a year) to find the daily amount, and subtracting it from the return itself. Using these new, after expenses, returns is possible to carry out the same regressions explained in the sections before to assess whether this procedure led to the same or different outcomes.

5 Results

In this section, the results of the analysis will be presented split into the four distinct hypotheses generated from the main research question in Section 2.4, to maintain coherence to the style employed up to now. The final part is occupied by the robustness checks that further corroborate the findings.

The first step, as delineated in the methods, is the creation of the performance metrics used in the analysis. Cumulative returns regarding the last 1-year and 6-month are computed for every ETF, and a summary table is displayed below:

Return	Obs.	Mean	St.Dev.	Median	Min.	Max.	Range
6-month	280	18.39	5.50	18.53	3.20	44.91	41.70
1-year	280	15.61	10.95	14.29	-26.42	49.49	75.92

In this table are described the main statistical measures of the cross-section of ETF returns for two different time periods. All the statistics are in percentage terms.

Table 5: Cumulative returns summary statistics

At first glance, these results may seem strange, however given the market instabilities during the last part of 2023, and some sectors (such as the energetic one) displaying huge losses during the first half of the trailing year, it is then logical to see higher returns in the shorter (thus more recent) period.

The second measure that will be employed, Sharpe ratios, are calculated starting from the previously mentioned cumulative returns, hence generating Sharpe ratios for both 6-month and 1-year periods. After their creation, these are tested by means of the Jobson-Korkie test statistic (Equation 3) to assess whether these values are significantly different from zero. Regarding 6-month Sharpe ratios, only around 1.5% of the values are not significant at a 5% significance level, number that increases to approximately 5% for the 1-year returns, with the majority of observation being significant at a 99.9% confidence interval. This tells us that these very significant values are not a consequence of sample noise, thereby successfully reflecting underlying effects.

5.1 Hypothesis 1: Sustainability correlation

ETF profitability is not significantly correlated with the magnitude of sustainability scores.

After having created all the performance measures that will be used as dependent variables, the first four regressions employing the *Portfolio Corporate Sustainability Score* (one for each performance measure) are estimated. After testing for heteroskedasticity using White tests, a heteroskedastic behavior is not found, eventually leading to the usage of the basic standard errors. Table 6 exhibit the results:

	Dependent variable:			
	6mRet	1yRet	6mSR	1ySR
	(1)	(2)	(3)	(4)
PCSS	−0.487***	−0.060	−0.090***	−0.015
	(0.099)	(0.206)	(0.013)	(0.015)
Constant	28.540***	16.866***	3.775***	1.177***
	(2.094)	(4.343)	(0.265)	(0.313)
Observations	280	280	280	280
R^2	0.080	0.0003	0.156	0.004
Adjusted R^2	0.076	-0.003	0.153	0.0003
Residual Std. Error (df = 278)	5.290	10.973	0.671	0.791
F Statistic (df = 1; 278)	24.024***	0.085	51.396***	1.076

This table depicts the results of the regressions generated following equation 4, where the four performance measures are regressed on PCSS, the *Portfolio Corporate Sustainability Score*. All returns are in percentage terms. Standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 6: PCSS regression table

Some general considerations can be made about the results presented above: the constant represents the performance value for a fully sustainable ETF (therefore yielding a PCSS of 0), while the negative PCSS coefficient depicts the inverse relation between the portfolio risk score and performance, where a higher PCSS value eventually leads to a worse outcome. However, this is significant only for the 6-month metrics, probably due to the aforementioned instability in the first half of the trailing year. As stressed also before, these analyses are not aimed at finding the determinants of returns and Sharpe ratios (as reflected by the very low adjusted R^2), but instead are focused on assessing whether sustainability factors have an impact on cross-sectional performance.

In the following regressions, the "current" PCSS (i.e. the one registered during the last month in the sample) is substituted by a lagged PCSS value. Given the periods of unequal length, different lags are needed for each of the four regressions: in Table 7 the four dependent variables are paired with their corresponding regressor. White tests are once again rejected in favor of classical standard errors.

	Dependent variable:			
	6mRet	1yRet	6mSR	1ySR
	(1)	(2)	(3)	(4)
PCSS (6m lag)	-0.512^{***}		-0.093^{***}	
	(0.100)		(0.013)	
PCSS (1y lag)		-0.121		-0.020
		(0.204)		(0.015)
Constant	29.185^{***}	18.163^{***}	3.863^{***}	1.276^{***}
	(2.132)	(4.360)	(0.270)	(0.314)
Observations	280	280	280	280
R^2	0.086	0.001	0.164	0.007
Adjusted R^2	0.083	-0.002	0.161	0.003
Residual Std. Error (df = 278)	5.271	10.968	0.667	0.790
F Statistic (df = 1; 278)	26.202^{***}	0.350	54.412^{***}	1.830

This table depicts the results of the regressions generated following equation 5, where the four performance measures are regressed on lagged values of PCSS, the *Portfolio Corporate Sustainability Score*. All returns are in percentage terms. Standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 7: Lagged PCSS regression table

When comparing these regressions with the previous ones, the same patterns of significance can be seen, along with very similar coefficient values. Consequently, it can be inferred that changes in sustainability ratings are very slow and unlikely to happen in short timespans, ultimately conferring robustness to Morningstar's methodology for constructing valid and complete PCSS. Given the strong ties with Table 6, the same conclusions can be drawn.

As expressed by Equation 6, the next step entails the usage of HCSS, the *Historical Corporate Sustainability Score*, whose value is computed using a weighted average of the previous 12 months' scores. As HCSS is calculated over a longer timespan, it also reflects recent changes in the ESG profile of an ETF, ultimately yielding more information than PCSS on its own. Testing for heteroskedasticity led to non-rejection even in this case. The regression coefficients are displayed in Table 8:

	Dependent variable:			
	6mRet	1yRet	6mSR	1ySR
	(1)	(2)	(3)	(4)
HCSS	-0.507^{***}	-0.093	-0.094^{***}	-0.018
	(0.099)	(0.206)	(0.013)	(0.015)
Constant	29.044^{***}	17.575^{***}	3.868^{***}	1.235^{***}
	(2.109)	(4.389)	(0.266)	(0.316)
Observations	280	280	280	280
R^2	0.086	0.001	0.168	0.005
Adjusted R^2	0.082	-0.003	0.165	0.002
Residual Std. Error (df = 278)	5.272	10.970	0.666	0.791
F Statistic (df = 1; 278)	26.080^{***}	0.205	56.217^{***}	1.472

This table depicts the results of the regressions generated following equation 6, where the four performance measures are regressed on HCSS, the *Historical Corporate Sustainability Score*. All returns are in percentage terms. Standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 8: HCSS regression table

These regressions share multiple results with the previous steps, with the coefficients being mostly between the ones in Table 6 and 7 probably as a consequence of the technique used to build this metric, that considers both recent and past scores. Given these similarities with the precedent tables, the findings derived up to now are strengthened even more.

The last batch of regressions related to this hypothesis tries to sum everything done up to now in a single model containing both the PCSS and the HCSS. By merging the two metrics, some valuable insights can be drawn to understand even more the dynamics of influence among the two regressors and the performance measures through the investigation of which one is the main driver and how the coefficients change when incorporating both together. As in the previous regressions, White tests for heteroskedasticity were used and non-rejected. Table 9 contains the results of the four comprehensive regressions.
Here, the situation is entirely different from the one presented before: both PCSS and HCSS register extreme values, with the former being highly positive as opposed to the latter extremely negative instead. In absolute terms, however, the HCSS values are higher, with the sum of the two effects being negative in case of equal (unlikely) or very similar values (much more probable) for both measures. It is also worth noting that all coefficients are significant, highlighting an improved model compared to the previous ones.

	Dependent variable:			
	6mRet	1yRet	6mSR	1ySR
	(1)	(2)	(3)	(4)
HCSS	−2.861**	−4.651*	−0.529***	−0.372**
	(1.166)	(2.429)	(0.146)	(0.175)
PCSS	2.356**	4.560*	0.435***	0.354**
	(1.163)	(2.422)	(0.146)	(0.174)
Constant	29.403***	18.270***	3.935***	1.289***
	(2.105)	(4.384)	(0.264)	(0.316)
Observations	280	280	280	280
R^2	0.099	0.013	0.194	0.020
Adjusted R^2	0.093	0.006	0.188	0.013
Residual Std. Error (df = 277)	5.243	10.921	0.656	0.786
F Statistic (df = 2; 277)	15.238***	1.876	33.388***	2.809*

This table depicts the results of the regressions generated following equation 7, where the four performance measures are regressed on HCSS and PCSS, the *Historical Corporate Sustainability Score* and the *Portfolio Corporate Sustainability Score*. All returns are in percentage terms. Standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 9: HCSS and PCSS regression table

In conclusion, a negative relation between these scores and the four performance measures can be clearly outlined from the regression coefficients. However, by looking at the 1-year metrics, it is also true that this link can be voided by periods of market instability, partly validating the hypothesis.

5.2 Hypothesis 2: E, S, and G impact

Governance drives the ETF performance more than Environmental and Social scores.

To study whether any of the single ESG risk scores is influencing the performance more than the others, the PCSS must be split into its components: *Environmental*, *Social*, *Governance*, and *Unallocated Risk Scores*. These three pillars (plus the last one for the mixed risk factors) are then used as regressors in the following models, with the aim of estimating the coefficients and their significance. As opposed to before, heteroskedasticity is detected in all four regressions, leading to the employment of heteroskedasticity-consistent (HC) standard errors (i.e. Eicker–Huber–White standard errors) to allow the fitting of these models containing heteroskedastic residuals. Table 10 displays the results of the four regressions:

	Dependent variable:			
	6mRet	1yRet	6mSR	1ySR
	(1)	(2)	(3)	(4)
Environmental	−0.838***	−0.513*	−0.134***	−0.054**
	(0.136)	(0.264)	(0.016)	(0.017)
Social	−0.594	−1.387**	0.052	−0.051
	(0.364)	(0.699)	(0.050)	(0.049)
Governance	0.925*	4.527***	−0.107	0.266***
	(0.525)	(0.980)	(0.069)	(0.067)
Unallocated	0.010	0.521	−0.095***	0.006
	(0.202)	(0.319)	(0.021)	(0.019)
Constant	21.008***	−1.642	2.912***	−0.274
	(3.443)	(6.746)	(0.466)	(0.436)
Observations	280	280	280	280
R^2	0.161	0.114	0.264	0.122
Adjusted R^2	0.148	0.101	0.253	0.110
Residual Std. Error (df = 275)	5.080	10.388	0.630	0.747
F Statistic (df = 4; 275)	13.156***	8.821***	24.640***	9.597***

This table depicts the results of the regressions generated following equation 8, where the four performance measures are regressed on *Environmental*, *Social*, *Governmental*, and *Unallocated Risk Scores*. All returns are in percentage terms. Heteroskedasticity-consistent (HC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 10: ESG regression table

The significance of these coefficients is not constant throughout the four different performance measures, and the most consistent out of the four scores is the *Environmental Risk Score*, achieving significance in all four regressions. Given the negative sign, an increase in the ETF's environmental footprint (i.e. an increment in the *Environmental Risk Score*) would lead to lower normal and risk-adjusted returns. Social and Unallocated Scores reveal inconsistent behavior across the different performance measures. Interestingly, the *Governmental Risk Factor* displays some significant positive coefficients, hinting at a positive correlation between Governmental Risk and returns (i.e. increase of returns for ETF displaying worse Governmental conditions).

Following the same line of reasoning used before for the PCSS, past values of E, S, G, and U are used to improve and verify further the results from Table 10. To account for the different periods of the metrics, two lags are used: the 6-month and the 1-year ones. Also in this case the White test is rejected in favor of heteroskedasticity, and Eicker–Huber–White standard errors are used to correct for it.

	Dependent variable:			
	6mRet	1yRet	6mSR	1ySR
	(1)	(2)	(3)	(4)
Environmental (6m lag)	−0.851***		−0.135***	
	(0.134)		(0.016)	
Social (6m lag)	−0.681*		0.036	
	(0.369)		(0.052)	
Governance (6m lag)	1.023**		−0.103	
	(0.520)		(0.066)	
Unallocated (6m lag)	−0.032		−0.098***	
	(0.178)		(0.021)	
Environmental (1y lag)		−0.644**		−0.063***
		(0.295)		(0.019)
Social (1y lag)		−1.129		−0.032
		(0.719)		(0.051)
Governance (1y lag)		4.071***		0.231***
		(1.018)		(0.071)
Unallocated (1y lag)		0.505*		0.008
		(0.293)		(0.019)
Constant	21.200***	−0.644	3.038***	−0.185
	(3.445)	(6.913)	(0.475)	(0.457)
Observations	280	280	280	280
R^2	0.173	0.109	0.259	0.119
Adjusted R^2	0.161	0.096	0.248	0.106
Residual Std. Error (df = 275)	5.043	10.418	0.632	0.748
F Statistic (df = 4; 275)	14.346***	8.368***	23.980***	9.286***

This table depicts the results of the regressions generated following equation 9, where the four performance measures are regressed on lagged values of *Environmental*, *Social*, *Governmental*, and *Unallocated Risk Scores*. All returns are in percentage terms. Heteroskedasticity-consistent (HC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 11: Lagged ESG regression table

A similarity with the previous regressions can be easily detected, as the significance and the coefficients remain more or less unaltered, adding robustness to the earlier results and further adding evidence to the claim that scores hardly face big changes, and even in that case, it would take long timespans to change significantly. The *Environmental Risk Score* is still the most influential, followed by the Governmental one with its peculiar behavior, while Social and Unallocated are still hardly significant.

With the results of the two previous tables, it can be said that while *Environmental Risk Score* is always significant throughout the eight regressions, *Governance Risk Score* comes in close second given its significance in six out of the eight regression and the notably high values of some of its coefficients compared to the others. For these reasons, the second hypothesis can be considered partially true but with two major corrections: firstly, also the *Environmental Risk Score* plays a very important role in driving the returns, and secondly, the correlation of the *Governance Risk Score* does not have the expected sign and instead works against the other metrics.

5.3 Hypothesis 3: Rating-specific performance

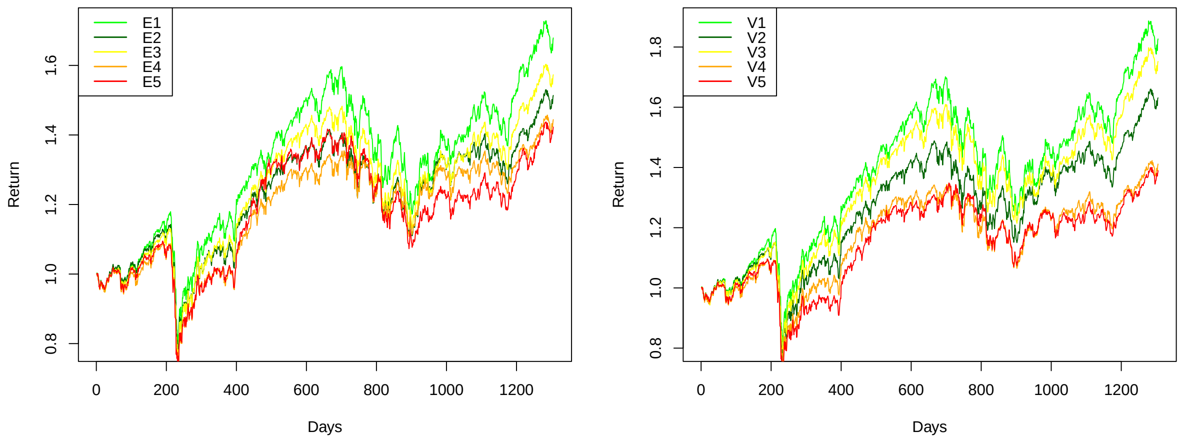
There is no significant overperformance of ETFs falling into higher sustainability brackets.

For the third hypothesis the focus is shifted to another measure, the *Morningstar Sustainability Rating*, derived from the comparison of the measures employed before with the the peer group. Since this measure is a *Rating*, the ETFs are split into five brackets, each one for a different ESG profile, from the worst (*Low*) to the best (*High*). This can be used to divide the ETFs into five distinct portfolios:

- P1 = *High*
- P2 = *Above Average*
- P3 = *Average*
- P4 = *Below Average*
- P5 = *Low*

These portfolios will be used in the upcoming analysis to assess whether the sustainable ones (i.e. P1 and P2) are subject to abnormal returns, thus making them overperform compared to the ETFs with worse ESG profiles (i.e. P4 and P5). After the split, a method is needed to compute daily portfolio returns over the whole period, and for this objective, two are the main possibilities: equally-weighted or value-weighted returns. By employing both, I get the possibility to explore the similarities and differences that these methods generate. For naming reasons, from this point onwards, the equally-weighted portfolios will be indicated with the letter "E", while value-weighted ones will use the letter "V". Both will also be followed by the corresponding number (e.g. E3 corresponds to the *Average* portfolio in its equally-weighted variant).

To start the examination of these portfolios, plotting cumulative returns over the time period of interest is a good way to set a baseline for further analysis, offering an intuitive and straightforward visual representation of the performance over time to have a better understanding of the different portfolio behaviors. It allows for a quick comparison (even though not a risk-adjusted one) to see which portfolio has generally performed better or worse throughout the period.



(a) Equally-weighted portfolios (b) Value-weighted portfolios

Figure 1: Cumulative returns for both weighting methods

Figure 1a and 1b display a general upward trend interrupted by two major collapses, the COVID-19 pandemic (the first, steeper one), and the market downturn of 2022 (the second one). By comparing the portfolios it can be clearly seen that throughout the 5-year period, the portfolio subjected to the highest returns for both EW and VW alternatives is the most sustainable one, while consistently worse performances are achieved by the *Below Average* and *Low* portfolios. This is emphasized even more for VW portfolios, where the difference with the best performer at the end is much bigger (around 45% instead of the EW 25%). However, these are only preliminary findings as they do not consider the risk profile of the portfolios, hence the improved performance can also be due to a higher risk exposure.

After having considered the normal returns, now it is time to factor in the riskiness of the asset and the exposure to anomalies to have a clear understanding of the pricing dynamics. Following the methodology delineated in Section 4.3, the first factor model to be estimated is the Capital Asset Pricing Model by Sharpe (1964). Tests for heteroskedasticity (White test) and autocorrelation (Breusch-Godfrey test) are rejected, leading to the adoption of Newey–West standard errors (also called HAC, heteroskedasticity and autocorrelation consistent standard errors). The results relative to the equally-weighted and value-weighted portfolios are contained in Tables 12 and 13:

	Dependent variable:				
	R_{E1} - R_f	R_{E2} - R_f	R_{E3} - R_f	R_{E4} - R_f	R_{E5} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.788***	0.796***	0.795***	0.789***	0.766***
	(0.021)	(0.020)	(0.022)	(0.025)	(0.023)
α	0.018	0.009	0.012	0.006	0.004
	(0.012)	(0.009)	(0.010)	(0.010)	(0.012)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.775	0.852	0.830	0.829	0.787
Adjusted R^2	0.775	0.851	0.830	0.829	0.787
Res. Std. Error (df = 1303)	0.515	0.403	0.437	0.434	0.484
F Statistic (df = 1; 1303)	4,493.723***	7,475.508***	6,352.802***	6,323.100***	4,808.447***

This table depicts the results of the regressions generated following equation 10, where the five equally-weighted portfolio returns are regressed on the market risk premium (R_M - R_f). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 12: EW CAPM regression table

	Dependent variable:				
	R_{V1} - R_f	R_{V2} - R_f	R_{V3} - R_f	R_{V4} - R_f	R_{V5} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.807***	0.828***	0.793***	0.801***	0.776***
	(0.022)	(0.019)	(0.023)	(0.022)	(0.027)
α	0.024*	0.014	0.021	0.004	0.003
	(0.014)	(0.010)	(0.013)	(0.010)	(0.011)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.746	0.842	0.766	0.842	0.796
Adjusted R^2	0.746	0.842	0.766	0.842	0.796
Res. Std. Error (df = 1303)	0.570	0.434	0.531	0.420	0.477
F Statistic (df = 1; 1303)	3,836.069***	6,967.214***	4,276.885***	6,948.022***	5,079.579***

This table depicts the results of the regressions generated following equation 10, where the five value-weighted portfolio returns are regressed on the market risk premium (R_M - R_f). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 13: VW CAPM regression table

Using this first (and very simple) model, despite α being slightly bigger for more sustainable portfolios (E1 - E5 is 0.014%), no significant abnormal returns can be observed for the equally-weighted portfolios. The value-weighted counterpart, instead, displays a significant α at 10% for the most sustainable portfolio (i.e. V1), compared to the other four still nonsignificant. Even though this factor model is the simplest out of all three, it still hints at a possible out-performance given by abnormal returns unpriced by the market. Regarding the two market risk premia, we can observe great significance, with an average β around 0.8 but with no clear pattern of change across our portfolios of interest.

The next model, the one theorized by Carhart (1997) 4-factor model, besides the market risk premium, uses three different market anomalies to price the asset in a more accurate way by considering three anomalies whose effect is not priced by the market. White tests for heteroskedasticity are again rejected along with Breusch-Pagan tests for error correlation, forcing the usage of heteroskedastic and autocorrelation consistent (HAC) standard errors once more. The regression coefficients for EW and VW portfolios using this model are in Tables 14 and 15:

	Dependent variable:				
	R_{E1} - R_f	R_{E2} - R_f	R_{E3} - R_f	R_{E4} - R_f	R_{E5} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.777***	0.764***	0.780***	0.762***	0.754***
	(0.024)	(0.021)	(0.022)	(0.022)	(0.022)
SMB	−0.037	−0.208***	−0.080**	−0.185***	−0.061
	(0.042)	(0.030)	(0.034)	(0.040)	(0.059)
HML	−0.173***	−0.005	−0.051**	0.078***	0.164***
	(0.029)	(0.023)	(0.024)	(0.024)	(0.031)
UMD	−0.072***	−0.037**	−0.040**	−0.006	0.012
	(0.024)	(0.018)	(0.020)	(0.019)	(0.023)
α	0.022*	0.009	0.013	0.004	0.002
	(0.012)	(0.009)	(0.011)	(0.010)	(0.012)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.786	0.858	0.832	0.838	0.799
Adjusted R^2	0.785	0.858	0.831	0.837	0.798
Res. Std. Error (df = 1300)	0.503	0.394	0.435	0.423	0.470
F Statistic (df = 4; 1300)	1,192.147***	1,971.747***	1,606.924***	1,680.228***	1,289.853***

This table depicts the results of the regressions generated following equation 11, where the five equally-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and three anomalies: size (SMB), value (HML), and momentum (UMD). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 14: EW Carhart 4-factor regression table

	Dependent variable:				
	R_{V1} - R_f	R_{V2} - R_f	R_{V3} - R_f	R_{V4} - R_f	R_{V5} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.796^{***}	0.793^{***}	0.778^{***}	0.778^{***}	0.734^{***}
	(0.024)	(0.019)	(0.023)	(0.021)	(0.024)
SMB	-0.083^{*}	-0.225^{***}	-0.102^{**}	-0.177^{***}	-0.236^{***}
	(0.044)	(0.032)	(0.044)	(0.035)	(0.048)
HML	-0.244^{***}	-0.052^{**}	-0.143^{***}	0.025	0.199^{***}
	(0.029)	(0.024)	(0.027)	(0.023)	(0.030)
UMD	-0.059^{**}	-0.053^{***}	-0.046^{*}	−0.003	−0.028
	(0.025)	(0.020)	(0.023)	(0.018)	(0.022)
α	0.028^{**}	0.015	0.023^{*}	0.002	0.0002
	(0.013)	(0.010)	(0.013)	(0.010)	(0.010)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.765	0.850	0.774	0.847	0.826
Adjusted R^2	0.764	0.849	0.773	0.847	0.826
Res. Std. Error (df = 1300)	0.550	0.424	0.523	0.414	0.440
F Statistic (df = 4; 1300)	$1{,}055.872^{***}$	$1{,}840.804^{***}$	$1{,}110.366^{***}$	$1{,}801.347^{***}$	$1{,}545.591^{***}$

This table depicts the results of the regressions generated following equation 11, where the five value-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and three anomalies: size (SMB), value (HML), and momentum (UMD). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 15: VW Carhart 4-factor regression table

Surprisingly, most of the exposures to the anomalies are negative for both weighting methods, indicating a negative loading on those factors, while the market premium remains the most influential regressor (although with a slightly lower magnitude). Confirming what was said before, the EW portfolio with the best ESG exposure registered some abnormal returns at the 10% significance level, indicating a modest outperformance compared to the others, with E4 and E5 registering values extremely lower compared to the first three portfolios. The conclusions that can be drawn for the VW ones are very similar to the previous, but an increase in the significance of abnormal returns stressed even more the overperformance of the *High* bracket compared to the others, although followed by V3, now significant at the 10% level.

The third and last model employed, the Fama and French (2015) 5-factor model, is the one containing the most factors, as apart from the basic market risk premia, size, and value, two more anomalies are used to price the ETFs: profitability (RMW) and investment (CMA). Also these regressions, after careful testing, resulted in the need for robust standard errors, once again opting for Newey-West ones. Equally- and value-weighted portfolios are used to determine the coefficients in Table 16 and 17:

	Dependent variable:				
	R_{E1} - R_f	R_{E2} - R_f	R_{E3} - R_f	R_{E4} - R_f	R_{E5} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.783***	0.771***	0.786***	0.777***	0.759***
	(0.025)	(0.022)	(0.024)	(0.024)	(0.024)
SMB	−0.054	−0.219***	−0.087**	−0.165***	−0.058
	(0.042)	(0.033)	(0.035)	(0.039)	(0.055)
HML	−0.132***	−0.010	−0.035	0.035	0.122***
	(0.045)	(0.035)	(0.038)	(0.037)	(0.047)
RMW	−0.087	−0.139**	−0.048	−0.016	−0.038
	(0.077)	(0.057)	(0.068)	(0.071)	(0.068)
CMA	−0.087	−0.015	−0.021	0.132*	0.071
	(0.086)	(0.067)	(0.078)	(0.075)	(0.080)
α	0.019	0.009	0.012	0.005	0.004
	(0.012)	(0.009)	(0.010)	(0.010)	(0.011)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.784	0.859	0.831	0.839	0.799
Adjusted R^2	0.783	0.859	0.830	0.838	0.798
Res. Std. Error (df = 1299)	0.506	0.393	0.436	0.422	0.470
F Statistic (df = 5; 1299)	940.810***	1,583.986***	1,278.370***	1,353.275***	1,033.439***

This table depicts the results of the regressions generated following equation 12, where the five equally-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and four anomalies: size (SMB), value (HML), profitability (RMW), and investment (CMA). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 16: EW Fama and French 5-factor regression table

	Dependent variable:				
	R_{V1} - R_f	R_{V2} - R_f	R_{V3} - R_f	R_{V4} - R_f	R_{V5} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.800***	0.802***	0.787***	0.788***	0.762***
	(0.026)	(0.021)	(0.025)	(0.023)	(0.026)
SMB	−0.102**	−0.236***	−0.108**	−0.174***	−0.195***
	(0.043)	(0.033)	(0.044)	(0.035)	(0.047)
HML	−0.207***	−0.038	−0.139***	−0.021	0.138***
	(0.048)	(0.039)	(0.044)	(0.035)	(0.044)
RMW	−0.089	−0.123**	−0.084	−0.092	0.001
	(0.082)	(0.060)	(0.080)	(0.067)	(0.064)
CMA	−0.081	−0.030	−0.008	0.078	0.218***
	(0.092)	(0.074)	(0.088)	(0.069)	(0.081)
α	0.026**	0.014	0.022*	0.004	0.001
	(0.013)	(0.009)	(0.013)	(0.009)	(0.010)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.764	0.850	0.773	0.848	0.828
Adjusted R^2	0.763	0.849	0.772	0.848	0.828
Res. Std. Error (df = 1299)	0.551	0.425	0.524	0.413	0.438
F Statistic (df = 5; 1299)	839.156***	1,466.958***	884.432***	1,452.142***	1,252.179***

This table depicts the results of the regressions generated following equation 12, where the five value-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and four anomalies: size (SMB), value (HML), profitability (RMW), and investment (CMA). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * p ≤ 0.1, ** p ≤ 0.5, *** p ≤ 0.01.

Table 17: VW Fama and French 5-factor regression table

Even in this case, the regressions are modestly loaded negatively towards all anomalies, with the previously employed (i.e. SMB and HML) much more significant than the two new additions. The market risk premium is still the most influential regressor. Across the two tables, more sustainable portfolios register much higher α compared to the non-sustainable ones. In assessing abnormal returns significance, EW portfolios do not show any, while the VW counterparts display it for both V1 and V3 α, with the first one being significant at 5% while the second only at 10% level.

In conclusion, the third hypothesis is rejected, as the most sustainable portfolio throughout several different models showed signs of significant abnormal returns accounting for around 2.5 basis points per day (circa 6.5% per year). This trend however is not continued by the *Above Average* portfolios, which do not show any sign of overperformance.

5.4 Hypothesis 4: Region-specific performance

The ESG ETF performance is not significantly different among investment regions.

After having analyzed the performance across the five different brackets, the focus for the fourth hypothesis shifts to geography, to grasp any difference in regional performance. As established in Section 4.4, the macroregions in which the sample is subdivided are 5:

- PGl = Global markets
- PEm = Emerging markets
- PEu = European markets
- PNa = North American markets
- PAs = Asian markets

Given the significant overperformance of the *High* bracket displayed in the previous hypothesis, the logical conclusion would be to subdivide this group into the five macroregions and understand where the abnormal returns are mostly concentrated. This, however, comes with the problem of underrepresentation of some macroregions (such as Emerging and Asian markets), problem that can be solved by including also the *Above Average* class. By adjusting the sample in this way, the analysis could however lose some significance because of the inclusion of the second (nonsignificant) group. After the split, to compute returns, equally-weighted and value-weighted returns are used once more, along with the same naming scheme (Exx for equally-weighted and Vxx for value-weighted portfolios). As the portfolios represent different geographical macroregions, regional-specific factors are employed to better track local trends and enhance accuracy.

Starting once again with CAPM, the five different models are estimated with both EW and VW portfolios, always using Newey-West standard errors to correct for heteroskedasticity and error correlation, both found in the sample through testing. The results can be found in Tables 18 and 19.
In these two tables, comparable values are registered, with the Emerging bracket showcasing the highest β (around 0.95), as opposed to the Asian market registering interestingly low values (circa 0.47), demonstrating a much lower sensitivity compared to the whole Asian market portfolio. All the market returns are very significant, maintaining the same trend seen in the general sustainability analysis, and thus remaining a substantial regressor. Regarding abnormal returns, some discrepancies can be detected, as the value-weighted portfolios capture some

	Dependent variable:				
	R_{EGl} - R_f	R_{EEm} - R_f	R_{EEu} - R_f	R_{ENa} - R_f	R_{EAs} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.702***	0.942***	0.872***	0.538***	0.456***
	(0.038)	(0.036)	(0.019)	(0.054)	(0.037)
α	0.016	0.0004	0.004	0.023	0.019
	(0.012)	(0.016)	(0.012)	(0.014)	(0.017)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.543	0.570	0.845	0.385	0.230
Adjusted R^2	0.543	0.569	0.845	0.385	0.230
Res. Std. Error (df = 1303)	0.712	0.829	0.453	0.909	0.908
F Statistic (df = 1; 1303)	1,547.774***	1,725.115***	7,115.052***	816.397***	389.616***

This table depicts the results of the regressions generated following equation 10, where the five equally-weighted portfolio returns are regressed on the market risk premium (R_M - R_f). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 18: EW Regional CAPM regression table

	Dependent variable:				
	R_{VGl} - R_f	R_{VEm} - R_f	R_{VEu} - R_f	R_{VNa} - R_f	R_{VAs} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.748***	0.964***	0.898***	0.560***	0.486***
	(0.042)	(0.037)	(0.019)	(0.056)	(0.034)
α	0.016	−0.003	0.008	0.024*	0.018
	(0.012)	(0.013)	(0.012)	(0.014)	(0.017)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.556	0.600	0.848	0.395	0.258
Adjusted R^2	0.556	0.600	0.848	0.395	0.258
Res. Std. Error (df = 1303)	0.739	0.797	0.462	0.926	0.897
F Statistic (df = 1; 1303)	1,631.357***	1,955.647***	7,262.291***	852.186***	453.582***

This table depicts the results of the regressions generated following equation 10, where the five value-weighted portfolio returns are regressed on the market risk premium (R_M - R_f). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 19: VW Regional CAPM regression table

overperformance in the North American market at the 10% level, while the equally-weighted ones do not show any significance.

It is now the turn of multifactor models, starting with the Carhart (1997) 4-factor model. Robust (HAC) standard errors are employed again after the rejection of the tests for heteroskedasticity and error correlation. Tables 20 and 21 display the coefficients for these regressions.

	Dependent variable:				
	R_{EGl} - R_f	R_{EEm} - R_f	R_{EEu} - R_f	R_{ENa} - R_f	R_{EAs} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.791***	0.892***	0.810***	0.535***	0.443***
	(0.037)	(0.041)	(0.020)	(0.064)	(0.040)
SMB	0.587***	−0.029	−0.350***	0.056	−0.107
	(0.066)	(0.113)	(0.043)	(0.064)	(0.092)
HML	−0.023	−0.134*	0.068**	0.089**	−0.120***
	(0.036)	(0.074)	(0.030)	(0.037)	(0.035)
UMD	0.057**	0.005	−0.091***	−0.043	−0.042
	(0.029)	(0.071)	(0.024)	(0.027)	(0.039)
α	0.026**	0.001	0.004	0.025*	0.022
	(0.012)	(0.016)	(0.011)	(0.014)	(0.017)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.603	0.573	0.869	0.398	0.239
Adjusted R^2	0.602	0.572	0.869	0.396	0.237
Res. Std. Error (df = 1300)	0.664	0.827	0.417	0.900	0.904
F Statistic (df = 4; 1300)	494.581***	436.062***	2,158.548***	214.476***	102.034***

This table depicts the results of the regressions generated following equation 11, where the five equally-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and three anomalies: size (SMB), value (HML), and Momentum (UMD). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 20: EW Regional Carhart 4-factor regression table

	Dependent variable:				
	R_{VGl} - R_f	R_{VEm} - R_f	R_{VEu} - R_f	R_{VNa} - R_f	R_{VAs} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.837***	0.903***	0.832***	0.557***	0.475***
	(0.039)	(0.044)	(0.019)	(0.066)	(0.037)
SMB	0.615***	−0.053	−0.418***	0.032	−0.090
	(0.067)	(0.104)	(0.041)	(0.064)	(0.088)
HML	−0.044	−0.142**	0.043	0.059	−0.106***
	(0.037)	(0.072)	(0.030)	(0.038)	(0.034)
UMD	0.044	0.049	−0.078***	−0.042	−0.024
	(0.030)	(0.068)	(0.025)	(0.028)	(0.038)
α	0.026**	−0.004	0.007	0.026*	0.020
	(0.011)	(0.013)	(0.011)	(0.014)	(0.016)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.616	0.605	0.873	0.402	0.265
Adjusted R^2	0.615	0.604	0.873	0.400	0.262
Res. Std. Error (df = 1300)	0.687	0.793	0.422	0.923	0.894
F Statistic (df = 4; 1300)	522.029***	498.057***	2,237.034***	218.265***	117.027***

This table depicts the results of the regressions generated following equation 11, where the five value-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and three anomalies: size (SMB), value (HML), and Momentum (UMD). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 21: VW Regional Carhart 4-factor regression table

Using this model, the results are in line with the previous findings, both for equally-weighted and for value-weighted portfolios, with slightly more significant outcomes in the former. Regarding the three anomalies, a variety of different loadings across the five macroregions can be noted, with both positive and negative exposures to these. Two α are found to be consistently significant within the two different strategies, namely the ones relative to the Global and North American markets, with the former being significant at 5% while the latter at 10% level.

The last multifactor model, the Fama and French (2015) 5-factor model, is now employed to study the two different groups of portfolios. Even in this case, the White and Breusch-Pagan tests led to rejection and subsequent introduction of Newey-West HAC standard errors. The results of these regressions are presented in Tables 22 and 23.

	Dependent variable:				
	R_{EGl} - R_f	R_{EEm} - R_f	R_{EEu} - R_f	R_{ENa} - R_f	R_{EAs} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.825***	0.895***	0.816***	0.530***	0.423***
	(0.037)	(0.042)	(0.021)	(0.065)	(0.041)
SMB	0.683***	−0.052	−0.374***	0.039	−0.108
	(0.070)	(0.106)	(0.043)	(0.083)	(0.090)
HML	−0.210***	−0.246***	0.146***	0.162**	0.067
	(0.062)	(0.078)	(0.048)	(0.063)	(0.080)
RMW	0.200***	−0.239***	−0.119**	−0.050	0.141
	(0.068)	(0.077)	(0.060)	(0.078)	(0.165)
CMA	0.197**	0.067	−0.116	−0.132	−0.269*
	(0.099)	(0.069)	(0.072)	(0.084)	(0.160)
α	0.022*	0.011	0.001	0.027*	0.019
	(0.012)	(0.015)	(0.011)	(0.015)	(0.017)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.609	0.580	0.867	0.398	0.245
Adjusted R^2	0.608	0.578	0.866	0.396	0.242
Residual Std. Error (df = 1299)	0.659	0.821	0.421	0.900	0.901
F Statistic (df = 5; 1299)	405.375***	358.373***	1,689.008***	171.751***	84.408***

This table depicts the results of the regressions generated following equation 12, where the five equally-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and four anomalies: size (SMB), value (HML), profitability (RMW), and investment (CMA). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 22: EW Regional Fama and French 5-factor regression table

	Dependent variable:				
	R_{VGl} - R_f	R_{VEm} - R_f	R_{VEu} - R_f	R_{VNa} - R_f	R_{VAs} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.860***	0.907***	0.831***	0.548***	0.458***
	(0.040)	(0.046)	(0.020)	(0.068)	(0.039)
SMB	0.701***	−0.051	−0.452***	0.003	−0.092
	(0.071)	(0.099)	(0.042)	(0.084)	(0.086)
HML	−0.172***	−0.214***	0.145***	0.174***	0.047
	(0.060)	(0.077)	(0.048)	(0.062)	(0.082)
RMW	0.217***	−0.123	−0.103*	−0.049	0.121
	(0.070)	(0.076)	(0.061)	(0.081)	(0.161)
CMA	0.085	0.041	−0.160**	−0.203**	−0.216
	(0.100)	(0.069)	(0.072)	(0.086)	(0.165)
α	0.023**	0.003	0.003	0.030**	0.018
	(0.011)	(0.012)	(0.011)	(0.014)	(0.016)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.621	0.606	0.872	0.404	0.269
Adjusted R^2	0.619	0.605	0.872	0.402	0.266
Res. Std. Error (df = 1299)	0.684	0.792	0.424	0.921	0.891
F Statistic (df = 5; 1299)	425.107***	400.007***	1,772.171***	176.240***	95.681***

This table depicts the results of the regressions generated following equation 12, where the five value-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and four anomalies: size (SMB), value (HML), profitability (RMW), and investment (CMA). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table 23: VW Regional Fama and French 5-factor regression table

Also here, major similarities between the two weighting methods can be tracked, with some minor differences in significance, while still maintaining the market risk premia as the most influential regressor. Once again, some abnormal returns are detected in both EW and VW portfolios for the Global and North American markets, conferring robustness and validating previous results.

After having presented all the relevant findings, also the fourth hypothesis can be confidently rejected as the majority of results pointed toward a North American and Global overperformance, with the latter probably being influenced by the former given the substantial impact of North American trade volumes and its market size on the global scenario.

5.5 Robustness checks

In order to confer even more strength to the results described before, ETF fees are implemented into the daily returns. As described in Section 4.5, the *Annual Report Net Expense Ratio* is a data point easily downloadable from Morningstar, and can be used to fulfill our scope. This metric, however, presents a few NA values that were promptly substituted by the average of the other observations. After the reestimation of all the regressions (whose tables can be found in Appendix A), the implementation of fund costs confirmed the robustness of these analyses. Apart from a few minor changes, this addition maintained intact most of the coefficients and significance levels, proving that the results found in the sections before are not invalidated by the introduction of fees, shifting the analysis from a theoretical framework to a real-world implementation.

6 Discussion and Conclusion

The last section of the thesis aims to synthesize and discuss the key findings of the research. As stated in Section 2.4, the final objective of the paper is to answer the question *"Does the integration of environmental, social, and governance (ESG) criteria into European ETFs enhance financial performance relative to traditional ETFs?"*. After the creation of four hypotheses with the purpose of exploring the research question from different angles and perspectives, several different methods are used to study them extensively.

The first hypothesis, *"ETF profitability is not significantly correlated with the magnitude of sustainability scores"* may seem at first glance easily rejectable because of some very significant results (in particular the ones relative to the 6-month returns and Sharpe ratios) showing a negative correlation between performance and risk scores and indicating that an ETF whose score is being increased (probably due to greater investments in companies with bad ESG scores) will face lower returns in the near future. Some questions about the robustness of these results in harsher periods are however raised, given the insignificant results relating to the trailing year where higher market instability was experienced.

In the second hypothesis, the *Portfolio Corporate Sustainability Score* is dissected into its three pillars to grasp whether *"Governance drives the ETF performance more than Environmental and Social scores"*. Also in this case the hypothesis was revealed to be only partially true, as the most important (and significant) factor is the *Environmental Risk Score*, with the *Governance Risk Score* coming in second place. This last factor, however, does not behave as expected, as its coefficients display very positive values, meaning that "better" governance would result in a loss of performance. In economic terms, a lower governance risk score would imply more restrictions to be fulfilled (e.g. board diversity, stakeholder engagement, and many more) and consequently a reduced profitability given by the suboptimal conditions in which the company has to operate to satisfy the ESG standards.

The third hypothesis, stating *"There is no significant overperformance of ETFs falling into higher sustainability brackets"* is completely rejected due to the several significant α found in the single- and multifactor models (both in their equally-weighted and value-weighted variants) during the estimation of the *High* portfolio, while still yielding no significance for any of the other portfolios. By now placing our attention on the coefficients, the delta between the best and the worst portfolio oscillates between 1.4 and 2.6 basis points on a daily basis (roughly between 3.6% and 6.8% yearly), further indicating the important difference among the top and bottom portfolios.

The fourth (and last) hypothesis tries to assess the geographical differences of this phenomenon, affirming that the *"The ESG ETF performance is not significantly different among investment regions"*. Also in this case a firm rejection can be outlined, as the abnormal returns found before in this analysis are revealed only regarding the ETFs invested in the North American and Global markets, with the others showing no significant results. It must be evidenced that the US market size roughly accounts for a third of the whole global market, therefore the underpricing discovered in the latter may be caused by the major American influence.

In conclusion, after having considered all the presented results, a positive answer can be given to the research question, ultimately confirming a degree of overperformance dictated by the ESG profile of the ETF. While this thesis may have brought interesting insights regarding ESG ETFs, its limitations must be also mentioned in an effort to indicate possible future research topics. Being the study limited to European exchanges, there is much to explore regarding other continents, and further analysis could be carried out with more risk-adjusting techniques, such as different ratios or the Modigliani risk-adjusted performance (M^2). Even while concentrating on the same European sample, more investigations regarding other ETF aspects can be carried out such as, for example, the influence of ESG factors on portfolios of different-sized ETFs. Further research on this topic is therefore needed to improve our knowledge of sustainable ETFs.

References

Bhattacharya, U., Loos, B., Meyer, S., & Hackethal, A. (2016). Abusing ETFs. *Review of Finance*, 21(3), 1217–1250. https://doi.org/10.1093/rof/rfw041

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57-82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Derwall, J., Koedijk, K., & Ter Horst, J. (2011). A tale of values-driven and profit-seeking social investors. *Journal of Banking & Finance*, 35(8), 2137-2147. https://doi.org/10.1016/j.jbankfin.2011.01.009

Dolvin, S. D., Fulkerson, J. A., & Krukover, A. (2017). Do'Good Guys' Finish Last? The Relationship between Morningstar Sustainability Ratings and Mutual Fund Performance. *The Journal of Investing*, ESG Special Issue 2019, 28(2), 77-91. https://joi.pm-research.com/content/28/2/77

Dumitrescu, A., Järvinen, J., & Zakriya, M. (2022). Hidden Gem or Fool's Gold: Can passive ESG ETFs outperform the benchmarks?. *International Review of Financial Analysis*, 86, 102540.

ETFGI. (2024, June 19). *ETFGI Global ETF and ETP Directory, May 2024*. ETFGI LLP. https://etfgi.com/research/report/26785/etfgi-global-etf-and-etp-directory-may-2024

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383–417. https://doi.org/10.2307/2325486

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47 (2), 427–465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Fama, E. F.,& French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, 116(1), 1-22. https://doi.org/10.1016/j.jfineco.2014.10.010

Filbeck, A., Filbeck, G., & Zhao, X. (2019). Performance assessment of firms following Sustainalytics ESG principles. *The Journal of Investing*, ESG Special Issue 2019, 28(2), 7-20. https://doi.org/10.3905/joi.2019.28.2.007

Folger-Laronde, Z., Pashang, S., Feor, L., & ElAlfy, A. (2020). ESG ratings and financial performance of exchange-traded funds during the COVID-19 pandemic. *Journal of Sustainable Finance & Investment*, 12(2), 490–496. https://doi.org/10.1080/20430795.2020.1782814

French, K. R. (2024, June 1). *Kenneth R. French - Data Library*. Dartmouth.edu. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Jobson, J. D., & Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance*, 889-908. https://doi.org/10.2307/2327554

Kanamura, T. (2021). Risk mitigation and return resilience for high yield bond ETFs with ESG components. *Finance Research Letters*, 41, 101866. https://doi.org/10.1016/j.frl.2020.101866

Kanuri, S. (2020). Risk and return characteristics of environmental, social, and governance (ESG) equity ETFs. *The Journal of Index Investing*, 11(2), 66-75. https://doi.org/10.3905/jii.2020.1.092

Kanuri, S., & McLeod, R. W. (2016). Sustainable competitive advantage and stock performance: the case for wide moat stocks. *Applied Economics*, 48(52), 5117–5127. https://doi.org/10.1080/00036846.2016.1170938

Memmel, C. (2003). Performance hypothesis testing with the Sharpe ratio. *Finance Letters*, 1, 21-23.

Meziani, A. S. (2020). It is still not easy being green for exchange-traded funds. *The Journal of Index Investing*, 10(4), 6-23. https://doi.org/10.3905/jii.2020.1.084

Morningstar Research (2021). *Morningstar Sustainability Rating methodology*. Morningstar. https://www.morningstar.com/content/dam/marketing/shared/research/methodology/SustainabilityRatingMethodology_2021.pdf

Nguyen, H. (2023). COVID-19: Performance of ESG ETFs and, ESG ETFs vs. their declared indexes. *Applied Finance Letters*, 12(1), 33-43. https://ojs.aut.ac.nz/applied-finance-letters/article/view/647

Nofsinger, J., & Varma, A. (2014). Socially responsible funds and market crises. *Journal of banking & finance*, 48, 180-193. https://doi.org/10.1016/j.jbankfin.2013.12.016

Pástor, Ľ., & Vorsatz, M. B. (2020). Mutual fund performance and flows during the COVID-19 crisis. *The Review of Asset Pricing Studies*, 10(4), 791-833. https://doi.org/10.1093/rapstu/raaa015

Pavlova, I., & de Boyrie, M. E. (2022). ESG ETFs and the COVID-19 stock market crash of 2020: Did clean funds fare better?. *Finance Research Letters*, 44, 102051. https://doi.org/10.1016/j.frl.2021.102051

Polbennikov, S., Desclée, A., Dynkin, L., & Maitra, A. (2016). *ESG ratings and performance of corporate bonds*. The Journal of Fixed Income, 26(1), 21. https://doi.org/10.3905/jfi.2016.26.1.021

Rompotis, G. G. (2022). The ESG ETFs in the UK. *Journal of Asset Management*, 23(2), 114-129. https://doi.org/10.1057/s41260-021-00251-z

Rompotis, G. G. (2023). The Performance of ESG ETFs in the US. *Capital Markets Review*, 31(2), 89-101. https://doi.org/10.3905/jesg.2023.1.070

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3), 425-442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Wikipedia Contributors. (2024, April 18). *Exchange-traded fund*. Wikipedia; Wikimedia Foundation. https://it.wikipedia.org/wiki/Exchange-traded_fund

Wikipedia Contributors. (2024, June 12). *Environmental, social, and governance*. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Environmental,_social,_and_governance

A Appendix A

This appendix presents all the tables regarding the robustness checks conducted in Section 5.5. The tables are arranged as follows:

- Hypothesis 1 (Sections 4.1, 5.1): Tables from A1 to A4
- Hypothesis 2 (Sections 4.2, 5.2): Tables A5 and A6
- Hypothesis 3 (Sections 4.3, 5.3): Tables from A7 to A12
- Hypothesis 4 (Sections 4.4, 5.4): Tables from A13 to A18

	Dependent variable:			
	6mRet	1yRet	6mSR	1ySR
	(1)	(2)	(3)	(4)
PCSS	−0.475***	−0.061	−0.080***	−0.004
	(0.097)	(0.195)	(0.011)	(0.014)
Constant	25.104***	10.700***	3.179***	0.405
	(2.042)	(4.120)	(0.241)	(0.297)
Observations	280	280	280	280
R^2	0.080	0.0003	0.150	0.0002
Adjusted R^2	0.076	-0.003	0.147	-0.003
Residual Std. Error (df = 278)	5.160	10.408	0.610	0.750
F Statistic (df = 1; 278)	24.111***	0.096	49.055***	0.067

This table depicts the results of the regressions generated following equation 4, where the four performance measures are regressed on PCSS, the *Portfolio Corporate Sustainability Score*. All returns are in percentage terms. Standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A1: PCSS regression table (after expenses)

	Dependent variable:			
	6mRet	1yRet	6mSR	1ySR
	(1)	(2)	(3)	(4)
PCSS (6m lag)	−0.500***		−0.083***	
	(0.098)		(0.012)	
PCSS (1y lag)		−0.119		−0.006
		(0.193)		(0.014)
Constant	25.733***	11.948***	3.261***	0.460
	(2.079)	(4.136)	(0.245)	(0.298)
Observations	280	280	280	280
R^2	0.086	0.001	0.158	0.001
Adjusted R^2	0.083	-0.002	0.155	-0.003
Residual Std. Error (df = 278)	5.141	10.402	0.607	0.749
F Statistic (df = 1; 278)	26.290***	0.377	52.154***	0.197

This table depicts the results of the regressions generated following equation 5, where the four performance measures are regressed on lagged values of PCSS, the *Portfolio Corporate Sustainability Score*. All returns are in percentage terms. Standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A2: Lagged PCSS regression table (after expenses)

	Dependent variable:			
	6mRet	1yRet	6mSR	1ySR
	(1)	(2)	(3)	(4)
HCSS	−0.495***	−0.092	−0.083***	−0.006
	(0.097)	(0.196)	(0.011)	(0.014)
Constant	25.599***	11.379***	3.259***	0.446
	(2.057)	(4.163)	(0.242)	(0.300)
Observations	280	280	280	280
R^2	0.086	0.001	0.161	0.001
Adjusted R^2	0.083	-0.003	0.158	-0.003
Residual Std. Error (df = 278)	5.142	10.405	0.606	0.749
F Statistic (df = 1; 278)	26.186***	0.222	53.453***	0.155

This table depicts the results of the regressions generated following equation 6, where the four performance measures are regressed on HCSS, the *Historical Corporate Sustainability Score*. All returns are in percentage terms. Standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A3: HCSS regression table (after expenses)

	Dependent variable:			
	6mRet	1yRet	6mSR	1ySR
	(1)	(2)	(3)	(4)
HCSS	−2.809**	−4.448*	−0.455***	−0.266
	(1.137)	(2.304)	(0.133)	(0.166)
PCSS	2.315**	4.358*	0.372***	0.261
	(1.134)	(2.297)	(0.133)	(0.166)
Constant	25.952***	12.043***	3.316***	0.486
	(2.053)	(4.158)	(0.240)	(0.300)
Observations	280	280	280	280
R^2	0.100	0.014	0.184	0.009
Adjusted R^2	0.093	0.006	0.178	0.002
Residual Std. Error (df = 277)	5.113	10.357	0.599	0.748
F Statistic (df = 2; 277)	15.326***	1.912	31.305***	1.316

This table depicts the results of the regressions generated following equation 7, where the four performance measures are regressed on HCSS and PCSS, the *Historical Corporate Sustainability Score* and the *Portfolio Corporate Sustainability Score*. All returns are in percentage terms. Standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A4: HCSS and PCSS regression table (after expenses)

	Dependent variable:			
	6mRet	1yRet	6mSR	1ySR
	(1)	(2)	(3)	(4)
Environmental	−0.818***	−0.490*	−0.121***	−0.037**
	(0.133)	(0.251)	(0.015)	(0.016)
Social	−0.572	−1.302*	0.027	−0.083*
	(0.355)	(0.665)	(0.045)	(0.045)
Governance	0.893*	4.271***	−0.059	0.316***
	(0.512)	(0.928)	(0.062)	(0.060)
Unallocated	0.004	0.481	−0.077***	0.028*
	(0.197)	(0.303)	(0.018)	(0.017)
Constant	17.750***	−6.860	2.333***	−0.976**
	(3.369)	(6.412)	(0.416)	(0.386)
Observations	280	280	280	280
R^2	0.160	0.113	0.242	0.129
Adjusted R^2	0.148	0.100	0.231	0.116
Residual Std. Error (df = 275)	4.957	9.856	0.579	0.704
F Statistic (df = 4; 275)	13.080***	8.778***	21.948***	10.146***

This table depicts the results of the regressions generated following equation 8, where the four performance measures are regressed on *Environmental*, *Social*, *Governmental*, and *Unallocated Risk Scores*. All returns are in percentage terms. Heteroskedasticity-consistent (HC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A5: ESG regression table (after expenses)

	Dependent variable:			
	6mRet	1yRet	6mSR	1ySR
	(1)	(2)	(3)	(4)
Environmental (6m lag)	−0.830***		−0.121***	
	(0.131)		(0.015)	
Social (6m lag)	−0.656*		0.012	
	(0.361)		(0.047)	
Governance (6m lag)	0.987*		−0.054	
	(0.507)		(0.060)	
Unallocated (6m lag)	−0.037		−0.079***	
	(0.174)		(0.019)	
Environmental (1y lag)		−0.614**		−0.045**
		(0.281)		(0.018)
Social (1y lag)		−1.058		−0.057
		(0.684)		(0.045)
Governance (1y lag)		3.837***		0.278***
		(0.963)		(0.064)
Unallocated (1y lag)		0.466*		0.031*
		(0.278)		(0.017)
Constant	17.938***	−5.893	2.436***	−0.942**
	(3.368)	(6.569)	(0.424)	(0.402)
Observations	280	280	280	280
R^2	0.172	0.108	0.241	0.120
Adjusted R^2	0.160	0.095	0.230	0.108
Residual Std. Error (df = 275)	4.922	9.884	0.580	0.707
F Statistic (df = 4; 275)	14.271***	8.333***	21.807***	9.405***

This table depicts the results of the regressions generated following equation 9, where the four performance measures are regressed on lagged values of *Environmental*, *Social*, *Governmental*, and *Unallocated Risk Scores*. All returns are in percentage terms. Heteroskedasticity-consistent (HC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A6: Lagged ESG regression table (after expenses)

	Dependent variable:				
	R_{E1} - R_f	R_{E2} - R_f	R_{E3} - R_f	R_{E4} - R_f	R_{E5} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.788***	0.796***	0.795***	0.789***	0.766***
	(0.021)	(0.020)	(0.022)	(0.025)	(0.023)
α	0.016	0.008	0.011	0.004	0.003
	(0.012)	(0.009)	(0.010)	(0.010)	(0.012)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.775	0.852	0.830	0.829	0.787
Adjusted R^2	0.775	0.851	0.830	0.829	0.787
Res. Std. Error (df = 1303)	0.515	0.403	0.437	0.434	0.484
F Statistic (df = 1; 1303)	4,493.723***	7,475.508***	6,352.802***	6,323.100***	4,808.447***

This table depicts the results of the regressions generated following equation 10, where the five equally-weighted portfolio returns are regressed on the market risk premium (R_M - R_f). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A7: EW CAPM regression table (after expenses)

	Dependent variable:				
	R_{V1} - R_f	R_{V2} - R_f	R_{V3} - R_f	R_{V4} - R_f	R_{V5} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.807***	0.828***	0.793***	0.801***	0.776***
	(0.022)	(0.019)	(0.023)	(0.022)	(0.027)
α	0.023*	0.013	0.020	0.003	0.001
	(0.014)	(0.010)	(0.013)	(0.010)	(0.011)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.746	0.842	0.766	0.842	0.796
Adjusted R^2	0.746	0.842	0.766	0.842	0.796
Res. Std. Error (df = 1303)	0.570	0.434	0.531	0.420	0.477
F Statistic (df = 1; 1303)	3,836.069***	6,967.214***	4,276.885***	6,948.022***	5,079.579***

This table depicts the results of the regressions generated following equation 10, where the five value-weighted portfolio returns are regressed on the market risk premium (R_M - R_f). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A8: VW CAPM regression table (after expenses)

	Dependent variable:				
	R_{E1} - R_f	R_{E2} - R_f	R_{E3} - R_f	R_{E4} - R_f	R_{E5} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.777***	0.764***	0.780***	0.762***	0.754***
	(0.024)	(0.021)	(0.022)	(0.022)	(0.022)
SMB	−0.037	−0.208***	−0.080**	−0.185***	−0.061
	(0.042)	(0.030)	(0.034)	(0.040)	(0.059)
HML	−0.173***	−0.005	−0.051**	0.078***	0.164***
	(0.029)	(0.023)	(0.024)	(0.024)	(0.031)
UMD	−0.072***	−0.037**	−0.040**	−0.006	0.012
	(0.024)	(0.018)	(0.020)	(0.019)	(0.023)
α	0.021*	0.008	0.012	0.003	0.001
	(0.012)	(0.009)	(0.011)	(0.010)	(0.012)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.786	0.858	0.832	0.838	0.799
Adjusted R^2	0.785	0.858	0.831	0.837	0.798
Res. Std. Error (df = 1300)	0.503	0.394	0.435	0.423	0.470
F Statistic (df = 4; 1300)	1,192.147***	1,971.747***	1,606.924***	1,680.228***	1,289.853***

This table depicts the results of the regressions generated following equation 11, where the five equally-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and three anomalies: size (SMB), value (HML), and Momentum (UMD). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * p ≤ 0.1, ** p ≤ 0.5, *** p ≤ 0.01.

Table A9: EW Carhart 4-factor regression table (after expenses)

	Dependent variable:				
	R_{V1} - R_f	R_{V2} - R_f	R_{V3} - R_f	R_{V4} - R_f	R_{V5} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.796***	0.793***	0.778***	0.778***	0.734***
	(0.024)	(0.019)	(0.023)	(0.021)	(0.024)
SMB	−0.083*	−0.225***	−0.102**	−0.177***	−0.236***
	(0.044)	(0.032)	(0.044)	(0.035)	(0.048)
HML	−0.244***	−0.052**	−0.143***	0.025	0.199***
	(0.029)	(0.024)	(0.027)	(0.023)	(0.030)
UMD	−0.059**	−0.053***	−0.046*	−0.003	−0.028
	(0.025)	(0.020)	(0.023)	(0.018)	(0.022)
α	0.027**	0.014	0.022*	0.001	−0.001
	(0.013)	(0.010)	(0.013)	(0.010)	(0.010)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.765	0.850	0.774	0.847	0.826
Adjusted R^2	0.764	0.849	0.773	0.847	0.826
Res. Std. Error (df = 1300)	0.550	0.424	0.523	0.414	0.440
F Statistic (df = 4; 1300)	1,055.872***	1,840.804***	1,110.366***	1,801.347***	1,545.591***

This table depicts the results of the regressions generated following equation 11, where the five value-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and three anomalies: size (SMB), value (HML), and Momentum (UMD). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A10: VW Carhart 4-factor regression table (after expenses)

	Dependent variable:				
	RF)	RF)	RF)	RF)	RF)
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.783***	0.771***	0.786***	0.777***	0.759***
	(0.025)	(0.022)	(0.024)	(0.024)	(0.024)
SMB	−0.054	−0.219***	−0.087**	−0.165***	−0.058
	(0.042)	(0.033)	(0.035)	(0.039)	(0.055)
HML	−0.132***	−0.010	−0.035	0.035	0.122***
	(0.045)	(0.035)	(0.038)	(0.037)	(0.047)
RMW	−0.087	−0.139**	−0.048	−0.016	−0.038
	(0.077)	(0.057)	(0.068)	(0.071)	(0.068)
CMA	−0.087	−0.015	−0.021	0.132*	0.071
	(0.086)	(0.067)	(0.078)	(0.075)	(0.080)
α	0.018	0.007	0.011	0.004	0.002
	(0.012)	(0.009)	(0.010)	(0.010)	(0.011)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.784	0.859	0.831	0.839	0.799
Adjusted R^2	0.783	0.859	0.830	0.838	0.798
Res. Std. Error (df = 1299)	0.506	0.393	0.436	0.422	0.470
F Statistic (df = 5; 1299)	940.810***	1,583.986***	1,278.370***	1,353.275***	1,033.439***

This table depicts the results of the regressions generated following equation 12, where the five equally-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and four anomalies: size (SMB), value (HML), profitability (RMW), and investment (CMA). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A11: EW Fama and French 5-factor regression table (after expenses)

	Dependent variable:				
	R_{V1} - R_f	R_{V2} - R_f	R_{V3} - R_f	R_{V4} - R_f	R_{V5} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.800***	0.802***	0.787***	0.788***	0.762***
	(0.026)	(0.021)	(0.025)	(0.023)	(0.026)
SMB	−0.102**	−0.236***	−0.108**	−0.174***	−0.195***
	(0.043)	(0.033)	(0.044)	(0.035)	(0.047)
HML	−0.207***	−0.038	−0.139***	−0.021	0.138***
	(0.048)	(0.039)	(0.044)	(0.035)	(0.044)
RMW	−0.089	−0.123**	−0.084	−0.092	0.001
	(0.082)	(0.060)	(0.080)	(0.067)	(0.064)
CMA	−0.081	−0.030	−0.008	0.078	0.218***
	(0.092)	(0.074)	(0.088)	(0.069)	(0.081)
α	0.025*	0.013	0.021	0.003	−0.0001
	(0.013)	(0.009)	(0.013)	(0.009)	(0.010)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.764	0.850	0.773	0.848	0.828
Adjusted R^2	0.763	0.849	0.772	0.848	0.828
Res. Std. Error (df = 1299)	0.551	0.425	0.524	0.413	0.438
F Statistic (df = 5; 1299)	839.156***	1,466.958***	884.432***	1,452.142***	1,252.179***

This table depicts the results of the regressions generated following equation 12, where the five value-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and four anomalies: size (SMB), value (HML), profitability (RMW), and investment (CMA). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A12: VW Fama and French 5-factor regression table (after expenses)

	Dependent variable:				
	R_{E1} - R_f	R_{E2} - R_f	R_{E3} - R_f	R_{E4} - R_f	R_{E5} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.702***	0.942***	0.872***	0.538***	0.456***
	(0.038)	(0.036)	(0.019)	(0.054)	(0.037)
α	0.015	−0.001	0.003	0.022	0.018
	(0.012)	(0.016)	(0.012)	(0.014)	(0.017)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.543	0.570	0.845	0.385	0.230
Adjusted R^2	0.543	0.569	0.845	0.385	0.230
Res. Std. Error (df = 1303)	0.712	0.829	0.453	0.909	0.908
F Statistic (df = 1; 1303)	1,547.774***	1,725.115***	7,115.052***	816.397***	389.616***

This table depicts the results of the regressions generated following equation 10, where the five equally-weighted portfolio returns are regressed on the market risk premium (R_M - R_f). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A13: EW Regional CAPM regression table (after expenses)

	Dependent variable:				
	R_{VGl} - R_f	R_{VEm} - R_f	R_{VEu} - R_f	R_{VNa} - R_f	R_{VAs} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.748***	0.964***	0.898***	0.560***	0.486***
	(0.042)	(0.037)	(0.019)	(0.056)	(0.034)
α	0.014	−0.004	0.007	0.023*	0.017
	(0.012)	(0.013)	(0.012)	(0.014)	(0.017)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.556	0.600	0.848	0.395	0.258
Adjusted R^2	0.556	0.600	0.848	0.395	0.258
Res. Std. Error (df = 1303)	0.739	0.797	0.462	0.926	0.897
F Statistic (df = 1; 1303)	1,631.357***	1,955.647***	7,262.291***	852.186***	453.582***

This table depicts the results of the regressions generated following equation 10, where the five value-weighted portfolio returns are regressed on the market risk premium (R_M - R_f). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A14: VW Regional CAPM regression table (after expenses)

	Dependent variable:				
	R_{EGl} - R_f	R_{EEm} - R_f	R_{EEu} - R_f	R_{ENa} - R_f	R_{EAs} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.791***	0.892***	0.810***	0.535***	0.443***
	(0.037)	(0.041)	(0.020)	(0.064)	(0.040)
SMB	0.587***	−0.029	−0.350***	0.056	−0.107
	(0.066)	(0.113)	(0.043)	(0.064)	(0.092)
HML	−0.023	−0.134*	0.068**	0.089**	−0.120***
	(0.036)	(0.074)	(0.030)	(0.037)	(0.035)
UMD	0.057**	0.005	−0.091***	−0.043	−0.042
	(0.029)	(0.071)	(0.024)	(0.027)	(0.039)
α	0.024**	−0.001	0.003	0.024*	0.021
	(0.012)	(0.016)	(0.011)	(0.014)	(0.017)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.603	0.573	0.869	0.398	0.239
Adjusted R^2	0.602	0.572	0.869	0.396	0.237
Res. Std. Error (df = 1300)	0.664	0.827	0.417	0.900	0.904
F Statistic (df = 4; 1300)	494.581***	436.062***	2,158.548***	214.476***	102.034***

This table depicts the results of the regressions generated following equation 11, where the five equally-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and three anomalies: size (SMB), value (HML), and Momentum (UMD). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A15: EW Regional Carhart 4-factor regression table (after expenses)

	Dependent variable:				
	R_{VGl} - R_f	R_{VEm} - R_f	R_{VEu} - R_f	R_{VNa} - R_f	R_{VAs} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.837***	0.903***	0.832***	0.557***	0.475***
	(0.039)	(0.044)	(0.019)	(0.066)	(0.037)
SMB	0.615***	−0.053	−0.418***	0.032	−0.090
	(0.067)	(0.104)	(0.041)	(0.064)	(0.088)
HML	−0.044	−0.142**	0.043	0.059	−0.106***
	(0.037)	(0.072)	(0.030)	(0.038)	(0.034)
UMD	0.044	0.049	−0.078***	−0.042	−0.024
	(0.030)	(0.068)	(0.025)	(0.028)	(0.038)
α	0.025**	−0.005	0.006	0.025*	0.019
	(0.011)	(0.013)	(0.011)	(0.014)	(0.016)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.616	0.605	0.873	0.402	0.265
Adjusted R^2	0.615	0.604	0.873	0.400	0.262
Res. Std. Error (df = 1300)	0.687	0.793	0.422	0.923	0.894
F Statistic (df = 4; 1300)	522.029***	498.057***	2,237.034***	218.265***	117.027***

This table depicts the results of the regressions generated following equation 11, where the five value-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and three anomalies: size (SMB), value (HML), and Momentum (UMD). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A16: VW Regional Carhart 4-factor regression table (after expenses)

	Dependent variable:				
	R_{EGl} - R_f	R_{EEm} - R_f	R_{EEu} - R_f	R_{ENa} - R_f	R_{EAs} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.825***	0.895***	0.816***	0.530***	0.423***
	(0.037)	(0.042)	(0.021)	(0.065)	(0.041)
SMB	0.683***	−0.052	−0.374***	0.039	−0.108
	(0.070)	(0.106)	(0.043)	(0.083)	(0.090)
HML	−0.210***	−0.246***	0.146***	0.162**	0.067
	(0.062)	(0.078)	(0.048)	(0.063)	(0.080)
RMW	0.200***	−0.239***	−0.119**	−0.050	0.141
	(0.068)	(0.077)	(0.060)	(0.078)	(0.165)
CMA	0.197**	0.067	−0.116	−0.132	−0.269*
	(0.099)	(0.069)	(0.072)	(0.084)	(0.160)
α	0.021*	0.009	−0.0004	0.027*	0.018
	(0.012)	(0.015)	(0.011)	(0.015)	(0.017)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.609	0.580	0.867	0.398	0.245
Adjusted R^2	0.608	0.578	0.866	0.396	0.242
Res. Std. Error (df = 1299)	0.659	0.821	0.421	0.900	0.901
F Statistic (df = 5; 1299)	405.375***	358.373***	1,689.008***	171.751***	84.408***

This table depicts the results of the regressions generated following equation 12, where the five equally-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and four anomalies: size (SMB), value (HML), profitability (RMW), and investment (CMA). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * p ≤ 0.1, ** p ≤ 0.5, *** p ≤ 0.01.

Table A17: EW Regional Fama and French 5-factor regression table (after expenses)

	Dependent variable:				
	R_{VGl} - R_f	R_{VEm} - R_f	R_{VEu} - R_f	R_{VNa} - R_f	R_{VAs} - R_f
	(1)	(2)	(3)	(4)	(5)
R_M - R_f	0.860***	0.907***	0.831***	0.548***	0.458***
	(0.040)	(0.046)	(0.020)	(0.068)	(0.039)
SMB	0.701***	−0.051	−0.452***	0.003	−0.092
	(0.071)	(0.099)	(0.042)	(0.084)	(0.086)
HML	−0.172***	−0.214***	0.145***	0.174***	0.047
	(0.060)	(0.077)	(0.048)	(0.062)	(0.082)
RMW	0.217***	−0.123	−0.103*	−0.049	0.121
	(0.070)	(0.076)	(0.061)	(0.081)	(0.161)
CMA	0.085	0.041	−0.160**	−0.203**	−0.216
	(0.100)	(0.069)	(0.072)	(0.086)	(0.165)
α	0.022*	0.002	0.003	0.029**	0.017
	(0.011)	(0.012)	(0.011)	(0.014)	(0.016)
Observations	1,305	1,305	1,305	1,305	1,305
R^2	0.621	0.606	0.872	0.404	0.269
Adjusted R^2	0.619	0.605	0.872	0.402	0.266
Res. Std. Error (df = 1299)	0.684	0.792	0.424	0.921	0.891
F Statistic (df = 5; 1299)	425.107***	400.007***	1,772.171***	176.240***	95.681***

This table depicts the results of the regressions generated following equation 12, where the five value-weighted portfolio returns are regressed on the market risk premium (R_M - R_f) and four anomalies: size (SMB), value (HML), profitability (RMW), and investment (CMA). All returns are in percentage terms. Heteroskedasticity and autocorrelation consistent (HAC) standard errors are presented in parentheses. Significance levels are the following: * $p \leq 0.1$, ** $p \leq 0.5$, *** $p \leq 0.01$.

Table A18: VW Regional Fama and French 5-factor regression table (after expenses)