Handelshøyskolen BI

GRA 19703 Master Thesis

Thesis Master of Science 100% - W

Predefinert informasjon

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

Deltaker

Navn:	Tobias Eskelund Johansen og Sander Djupvik

Informasjon fra deltaker

Tittel *:	Uncovering the Puzzle of ESG Score Adjustments: An Analysis of Systemic Impacts and Momentum Dynamics
Navn på veileder *:	Kjell Jørgensen

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

Gruppe

Gruppenavn:	(Anonymisert)
Gruppenummer:	326
Andre medlemmer i gruppen:	

Uncovering the Puzzle of ESG Score Adjustments: An Analysis of Systemic Impacts and Momentum Dynamics

Master Thesis

MSc in Finance

Supervisor:
Kjell Jørgensen

Oslo, July 3, 2023

ABSTRACT

In this research paper, we explore the presence of momentum in ESG scores with the possibilities of utilizing the potential momentum to generate alpha. Three drivers - company, industry, and country - are analyzed through regression analysis using MSCI ESG data to develop a systematic, investment, and predictive model. Our findings show that ESG score changes exhibit a short-term inverse relationship with preceding periods, but individual momentum becomes apparent over longer periods. We also identify company, industry, and country factors as drivers of momentum in ESG scores. In conclusion, our research indicates weak momentum tendencies in ESG scores

Acknowledgments

We wish to express our sincere gratitude towards our supervisor, Kjell Jørgensen, whose expertise and generous time allocation significantly enriched this paper's quality. We also extend our appreciation to MSCI ESG Research LLC and the BI Norwegian Business School. Their provision of crucial resources played an integral role in successfully completing our research.

Disclaimer

Contents

List of Figures

List of Tables

1 Introduction and motivation

In this paper, we investigate the puzzle of ESG score adjustments and research if there exists momentum in ESG scores, possibly utilizing the momentum to generate alpha. We believe that identifying "on the rise" companies early on will lead to investment opportunities in green stocks before climate risk premiums get priced in. This upward momentum might be due to company engagements by big institutional owners. Companies that get targeted are often seen as having high potential and a high willingness to improve (Dimson et al., 2015). This is relevant because identifying momentum might be identifying either engaged or self-engaging companies. In the paper they also find evidence of significant excess returns of 7% over the next year after a successful engagement addressing environmental, social, and governance concerns. It is not unlikely that successful engagements also result in an increased ESG score as the companies manage to improve or fix ESG-related issues. However, it is important to note that the relationship between company engagement and ESG scores is complex, and an improvement in a company's ESG score is not guaranteed because of a successful engagement. This topic is highly interesting due to the increased focus on ESG in investment decision-making. Going forward, ESG is likely to be even more incorporated, and therefore it is reasonable to assume that it will have an even bigger impact on stock returns. MSCI finds that "ESG investing is growing exponentially as more investors and issuers utilize ESG and climate data and tools to support their financial decision-making." (*ESG Investing*, n.d.). Additionally, momentum in ESG scores is little researched beforehand, and therefore we could provide new insight into ESG scores' influence on stock returns. Giese & Lee (2019) wrote a conclusive statement in their article from 2019: "There is some evidence that ESG momentum (changes in ESG characteristics) is linked with portfolio performance, but a longer time series is needed to verify the existence of an

ESG risk premium".

This paper tackles the challenge of proving trends in ESG scores and how to exploit them in relation to investment advantages. The motivation is therefore twofold. First, to address a crucial gap in the landscape of research regarding ESG momentum. Second, if we can predict up- and down-adjustments, research suggests that abnormal returns can be harvested through a strategy of buying up-adjustments and selling down-adjustments. Given the limited research on ESG momentum, we will apply a diversified set of models to gather robust evidence supporting or disregarding momentum in ESG scores. We aim to bridge knowledge gaps, contribute new insights, and shape future ESG-oriented investment strategies.

To provide a comprehensive analysis, we first employ regression analysis to uncover the underlying relationships between a firm's current ESG scores and their preceding scores. Then, we look at the effects industries and countries have on scores in fixed effect models. Building on the findings, the study delves deeper into these relationships using three distinct models. First, a ranking model is used to decode systematic relationships existing between scores over time. The second approach implements a stock-picking strategy model, aiming to extract investment value from our insights on ESG score progression. Lastly, we introduce a predictive model to forecast ESG score changes by identifying trends based solely on previous scores. We will use ESG data from MSCI and Refinitiv, stock price data from EOD Historical Data, factor returns from Kenneth French's website, ESG index data from MSCI, and bond data for risk-free rate from FRED.

Our results reveal an interesting dynamic of short-term score reversions. Finding significant changes in a company's ESG scores partially reverted in the subsequent year. This insight led to the creation of momentum portfolios, both positive and negative. However, these portfolios did not yield significant positive excess returns, indicating that short-term momentum effects alone might not be sufficient for alpha generation. Moreover, we observed potential momentum over extended periods, suggesting long-term trends might be more predictable.

2 Literature review

The literature on momentum in ESG scores is minimal, Sankar et al. (2019) published a report on ESG rating and momentum. They analyze if improvements in ESG rating generate excessive returns for investors. By sorting out the top 30% highest ESG scores from each industry semi-annually, they construct a positive and negative momentum portfolio of firms that experience an increase or decrease in ESG score equal to or greater than 10%. The result from the report, based on and compared to the European stock index STOXX600, is that the positive momentum portfolio outperforms the index by a cumulative return of 23.5% between March 2013 and January 2019. However, because of few instances where firms experienced a fall in scores greater than 10%, they merged the negative with the neutral portfolio, neglecting the negative momentum effect. This is something we are interested in investigating. If the negative momentum portfolio experiences negative returns, it might be a good idea to create a long/short strategy instead. Furthermore, they did not investigate how the ESG score evolved during the period of portfolio inclusion; instead, they exclusively focused on score changes from the previous period.

Another paper, written by Bekaert et al. (2023), creates two sector-neutral portfolios of the 10% highest and lowest based on absolute and relative ESG momentum, going long the top 10% and short the bottom 10%. Furthermore, they compare the portfolio to the MSCI US index and find that the relative momentum portfolio generates a significant alpha of 0.47% monthly. In the case of the absolute momentum portfolio, contrary to the relative momentum, it did not display significant alphas.

Others try to predict scores using machine learning. However, they mostly use alternative sources such as fundamental data like D'Amato et al. (2022), who found evidence that balance sheet data provide a crucial element in explaining

ESG scores.

A study by Shanaev & Ghimire (2022) claims to be the first to document the importance of Environmental, Social, and Governance (ESG) rating changes, rather than ESG rating levels, for stock performance. Applying a calendar-time portfolio methodology on US-traded firms rated by MSCI in 2016–2021, they found that while ESG rating upgrades are associated with small and sometimes insignificant positive abnormal returns, downgrades are consistently detrimental to stock performance and lead to statistically and economically significant negative abnormal returns of -1.0% to -1.4% per month. The effects are stronger in firms already considered ESG leaders, suggesting that institutional investors use best-in-class positive screening. During the COVID-19 period, ESG rating upgrades show a pronounced positive effect which can be explained by individual investors' increased use of ESG ratings.

2.1 Backfilling and ESG scores

A well-known problem in ESG scores is the "backfill problem" of scores, also known as "attenuation bias". Yahya & Vaihekoski (2021) describe backfill bias as the lack of reliability to historical data on ESG scores as the historical data is limited because companies may retroactively improve their ESG scores to appear more attractive to investors. This can lead to an overestimation of a company's true level of ESG performance and make it difficult to evaluate a company's commitment to sustainability accurately.

According to research conducted by Berg et al. (2022), frequent and consistent modifications in the ESG scores provided by Thomson Reuters Refinitiv were found, which further supports the attenuation bias problem. Therefore, investors should be cautious when using ESG scores as the sole metric for evaluating a company's commitment to sustainability. As a result, future studies should look into disentangling management quality and ESG performance by incorporating firm-fixed effects, according to Berg et al. (2022). At the same time, Yahya & Vaihekoski (2021) point out how a firm's ESG performance changes over time, including back-testing strategies to confirm the backfilling bias, resulting from unreliable historical data on ESG scores. Both of these studies will require a more extensive time- series than currently available.

2.2 Determinants of ESG rating

Exploring the factors that influence ESG ratings, Crespi & Migliavacca (2020) empirically analyzed firm, country, and time factors in the financial industry. Their main findings are linear growth in ESG scores, enhanced by size and profitability, which is heavily influenced by their operational country's economic and social development. However, ESG rating may also be critically influenced by agency challenges, given the providers and major rating agencies' business relations with many companies. A conflict of interest may inevitably arise and reduce the credibility of the ESG ratings.

Somewhat interesting is the finding of Crespi & Migliavacca (2020), pointing at the different patterns which the three pillars of E, S, and G follow. Unlike the environmental and social pillars, the governance pillar tends to react opposite to the firm and country factors. Additionally, the governance pillar seems to drive the overall increase in the ESG score over time.

Crespi & Migliavacca (2020) also present the argument for which the different trends and rationale are driven by larger companies primarily prioritizing environmental and social approaches to improve their ESG score. In contrast, smaller companies focus on their governance quality to improve ESG scores. The authors' primary rationale behind the divergence between larger and smaller, and less capitalized companies is that improving environmental and social pillars is more capital intensive compared to improving the governance pillar, which Crespi & Migliavacca (2020) argue may explain their focus on governance to improve ESG scores. Finally, the two authors reason that the cost of CSP could be seen in relation to the positive performance in the governance pillar's association with small financial firms with low common equity. The fact that the three ESG pillars follow different trends may impact the different findings in this study. Hence, we should keep it in the back of our minds while studying this subject and explaining the potential drivers of our findings.

3 Testable hypotheses

Are there systemic effects in changes in ESG scores? Is autocorrelation present, for example, when a company improves their ESG score – is this followed by more increases in the future? Can we identify trends in these adjustments? And if so, can we generate alpha using a momentum strategy?

We have theorized three crucial momentum drivers: the company itself, the industry it belongs to, and the country or region it is located in. At the company level, increasing demand from investors for heightened ESG standards stimulates businesses to undertake initiatives aimed at score improvement. Implementing these initiatives can be a gradual process, which means an initial upward adjustment may be trailed by additional enhancements as the process unfolds. Conversely, companies backed by investors who undervalue the significance of ESG may witness a decline in their score, especially as regulations tighten and these companies persist in neglecting the importance of ESG adherence, their scores will continue to fall.

At the **industry level**, implementing ESG initiatives by one company can lead to competitive pressure for others in the same industry to follow suit. This can create a ripple effect within an industry, promoting better ESG practices and performance to stay competitive. An additional factor here is that companies in the same industry are often under the same regulations and rules, which affects all firms equally.

At the **country/region level**, new regulations or taxes can be introduced, which forces companies to take actions that indirectly improve their ESG scores. For example, Sweden introduced emissions taxes in 1991 based on the amount of emission a company produced, encouraging companies to reduce their emissions to avoid higher taxation. According to the Swedish Environmental Protection Agency, the carbon tax has contributed to a reduction in emissions of approximately 27% between 1990 and 2018 (*Looking Back on 30 Years of Carbon Taxes in Sweden*, 2020). Countries can also introduce incentives promoting a greener business, aligning shareholder objective functions towards ESG (Bonham & Riggs-Cragun, 2022). This can lead to improvements in ESG scores for the entire country/region.

Null Hypothesis (H0): There exists no momentum in ESG scores -the changes in ESG scores are random and do not exhibit a trend over time. **Alternative Hypothesis (H1)**: There exists momentum in ESG scores -the changes in ESG scores demonstrate to some degree a consistent trend over time.

4 Methodology

4.1 Regression Analysis of the Data

Definition: In the regressions and models, we use the percentage change over the last n periods as independent variables. They are also lagged by one period to ensure we only use available data at that time. To make it easier to follow along, we define $\mathbf{n}\lambda_\mathbf{i}$ as this change over time where i is an indicator for quarterly (Q) or yearly (Y) and n is the number of periods change. Example $3\lambda_Y$ means the percentage change over the last 3 years.

4.1.1 Run Pooled Regression

In order to assess the possibilities for momentum in ESG scores, we need to assess whether there exists a significant relationship between changes in ESG scores and changes with various timeframes ($n\lambda$'s) in ESG scores, along with the direction of these coefficients. The fact that we are working with a panel data set makes it easier to handle large number of companies in a pooled regression, compared to many AR models for each company. Another argument favoring pooled regression is the incorporation of effects of different entities as industries and countries, which are not possible to incorporate into an AR model. Such a model will also allow for more flexibility in the structure of lags; hence we will be able to construct the λ's more appropriately. For example, we will be able to investigate 1-4λ_Q. While in an AR model we will need to construct four different models for each λ_Q.

We run the pooled regressions for quarterly and yearly frequency, where we set change in ESG scores as the dependent variable and $n\lambda_i$ as the explanatory variables to minimize the number of variables. We specify the use of relative changes (percentage changes) rather than absolute changes throughout this paper, as Bekaert et al. (2023) showed that relative ESG momentum portfolios have the best performance.

For the quarterly frequency, we test score changes for 1,2,3 and 4 quarters ($1\lambda_Q - 4\lambda_Q$) to catch short-term trends. For the yearly frequency, we choose score changes from 1 year up to 7 years ($1\lambda_Y - 7\lambda_Y$), in order to catch long-term trends. We do not assess the monthly frequency of ESG scores as they are usually not scored as frequently. Most ESG scores are scored on a quarterly or yearly basis. We run similar regressions for the three pillar scores E, S, and G as we did for the total ESG score to assess whether the pillars λ contain potentially different information than the λ for overall ESG scores. We do this due to Crespi & Migliavacca (2020) findings about the pillars following individual trends. Below we outline the two regression equations we run for each frequency separately.

Quarterly frequency:

$$y_{i,P} = \alpha + \beta_1 x_{i,1\lambda_{Q,P}} + \beta_2 x_{i,2\lambda_{Q,P}} + \beta_3 x_{i,3\lambda_{Q,P}} + \beta_4 x_{i,4\lambda_{Q,P}} + u_{i,t} \qquad (1)$$

Yearly frequency:

$$\begin{aligned} y_{i,P} = \alpha &+ \beta_1 x_{i,1\lambda_{Y,P}} + \beta_2 x_{i,2\lambda_{Y,P}} + \beta_3 x_{i,3\lambda_{Y,P}} + \beta_4 x_{i,4\lambda_{Y,P}} \\ &+ \beta_5 x_{i,5\lambda_{Y,P}} \beta_6 x_{i,6\lambda_{Y,P}} + \beta_7 x_{i,7\lambda_{Y,P}} + u_{i,t} \qquad (2) \end{aligned}$$

Where P *= ESG, E, S, G*

When running such a pooled regression, the implicit assumption is that there is homogeneity in both the time-dimension and the cross-sectional dimension, which in turn means that the average values of the variables and the relationship between them hold for all the data (Brooks, 2019). The belonging null and alternative hypothesis for the pooled regression is as follows:

$$H_0 : \beta_1 = ... = \beta_i = 0, \quad H_1 : \beta_1 \neq or \ ... \ or \ \beta_i \neq 0, \quad where\ i = 4\ and\ 7$$

Framed in words, the null hypothesis expresses that there is no relationship between the independent and dependent variables, meaning that λ'shas no significant explanatory effect on ESG scores. While the alternative hypothesis expresses a significant relationship between the independent variables and the dependent variable, hence the λ's has a significant explanatory effect on the current change in score. In order to evaluate the hypothesis, we compare the F-statistic of the regression toward the critical value at a significance level of 5%.

$$Reject\ H_0\ \ if \quad F - statistic > t_{crit}$$

$$Fail\ to\ Reject\ H_0\ \ if \quad F - statistic < t_{crit}$$

We also evaluate the p-values of each of the single independent variables to an alpha (significance level) of 5%, if $p - value < \alpha$, we can report significant λ's.

4.1.2 Fixed-Effect Regression

In chapter 2 we mentioned our theory about how industry and country may impact changes in scores and potentially be a momentum driver. Henceforth, we want to test for industry-fixed effects and country-fixed effects. In order to test for these effects we assign a dummy variable, a variable that takes the values of 0 or 1, to each company within these two categories in separate regressions. The dummies indicate the presence or absence of the categorical effect on each industry or country via the size of the dummy intercept. We also need to be aware of the potential issues of perfect multicollinearity between the dummy variables and the intercept, the "dummy variable trap". Hence, we leave out one of the industries or countries (Brooks, 2019). The effect of this industry or country will show up in the regression intercept. Such a fixed effects model will allow us to catch effects that differ between the various fixed variables and control for unobserved heterogeneity in the data.

The data set will be limited to a subperiod from 2013 and forward due to frequent backfilling before 2013, as outlined in subsection 2.1. Furthermore, we expect to see differences in the results between the two datasets from MSCI and Refinitiv due to the paper of Berg et al. (2019) findings of low correlation between different score providers of ESG scores, averaging only 0.54.

Since we are only testing for entity-fixed effects, we assume that no fixed time-varying factors influence the changes in ESG scores (Brooks, 2019). This assumption, however, might not always hold as dynamics change over time in industries and countries due to factors such as market competition, innovation, and countries' economic, political, and regulatory landscapes. An example of this is if a company is reclassified to a new industry. Also, while the model captures within-country changes, it does not account for between-country variations or any global trends that might affect all companies regardless of location. Such limitations must be carefully considered when interpreting this type of analysis results.

4.1.2.1 Industry Fixed Regression

Companies commonly engage in self-comparisons and benchmark themselves against their industry peers, which fosters industry-specific competitive dynamics, as highlighted by Smith et al. (2005). These dynamics can fuel innovation and enhance overall performance within the industry, including influencing changes in ESG scores. By fixing the industries, we can investigate how industries affect companies' changes in scores. For instance, some industries might have more heavy regulations towards the environment, influencing companies within these industries to improve their ESG scores; this will show up as a positive industry intercept.

It is reasonable to expect that changes in ESG scores may reflect the unique challenges and dynamics within different industries. For instance, industries heavily reliant on natural resources, like mining, oil, and gas, may face more environmental challenges. Any major improvements or deteriorations in their environmental practices could cause larger fluctuations in their ESG scores. On the other hand, industries more focused on technology and professional services may experience less variation in the environmental pillar due to the nature of their operations. However, they may see greater fluctuations in their governance and social scores due to factors such as shifts in data privacy regulations, changes in diversity and inclusion practices, or significant company-wide ethical or legal issues. The equation for the industry-fixed regression is as follows:

$$y_{n,P} = \alpha + \sum_{i=1}^{I} \beta_i x_{n,i\lambda_{C,P}} + \sum_{n=1}^{N} \mu_{\text{industry}} D\text{industry}_n + v_{n,t} \ , \tag{3}$$

$$I = 4, 7, \quad C = \text{Q,Y}, \quad P = \text{ESG, E, S, G}$$

Defining Equation 3 as the restricted regression, while the pooled regressions in Equation 1 and Equation 2 are considered the unrestricted regressions. Below, the belonging null and alternative hypothesis is displayed:

$$H_0 : \ \mu_1 = \ ... \ = \mu_{\text{industry}}, \quad H_1 : \ \mu_1 = \ ... \ \mu_{\text{industry}}$$

Translated into words, the null hypothesis states that there is homogeneity in the cross-section and no industry effects, whereas the alternative hypothesis states heterogeneity in the cross-section and, thus, industry effects. To evaluate the hypothesis, we compare the F-statistic with the critical value at a significance level of 5%; hence this will be the outcome with 95% certainty.

$$Reject\ H_0 \ \ if \quad F - statistic > t_{crit}$$

$$Fail\ to\ Reject\ H_0 \ \ if \quad F - statistic < t_{crit}$$

Furthermore, we focus on the industry intercepts we obtain from the model and evaluate the p-values to an alpha (significance level) of 5%, if $p - value < \alpha$, we can report significant industry intercepts.

4.1.2.2 Country-Fixed Regression

Companies are affected by factors such as regulatory environment, cultural differences, economic factors, and political stability within their country. Countries that apply stricter environmental regulations may encourage companies to improve their ESG score, which would be associated with a positive country intercept in a country-fixed regression. Countries experiencing reduced environmental regulations may be associated with negative changes in ESG scores reflected with a negative country intercept to show that the country has a negative influence on the changes in ESG scores.

It is reasonable to expect differences between countries due to cultural, social, and political factors to impact changes in ESG scores. Any major policy changes, enforcement of environmental regulations, or shifts in national sustainability initiatives could cause larger fluctuations in the ESG scores of companies operating in those countries. Conversely, countries strongly emphasizing corporate responsibility and governance might see more stability in their ESG scores. This could result from factors such as changes in national policies related to data privacy, alterations in regulations promoting diversity and inclusion, or significant changes in the country's political stability or corruption levels. We run the following country-fixed regression equation:

$$y_{n,P} = \alpha + \sum_{i=1}^{I} \beta_i x_{n,i\lambda_{C,P}} + \sum_{n=1}^{N} \mu_{\text{country}} D\text{country}_n + v_{n,t} \ , \qquad (4)$$

$$I = 4, 7, \quad C = \text{Q,Y}, \quad P = \text{ESG, E, S, G}$$

Equation 4 is the restricted regression, while the pooled regressions in Equation 1 and Equation 2 are the unrestricted regressions. The belonging null and alternative hypotheses are displayed below:

$$H_0: \ \mu_1 = \ ... \ = \mu_{\text{country}}, \quad H_1: \ \mu_1 = \ ... \ \mu_{\text{country}}$$

Translated into words, we have a hypothesis that states that there is homogeneity in the cross-section and no country effects. Meanwhile, the alternative hypothesis states heterogeneity in the cross-section and country effects. As we did for the industry-fixed regression, we compare the F-statistic towards the critical value at a significance level of 5%, meaning that it is 95% certainty the outcome.

$$Reject\ H_0\ \ if \quad ; F - statistic > t_{crit}$$

$$Fail\ to\ Reject\ H_0\ \ if \quad F - statistic < t_{crit}$$

We obtain the country intercept from the regression model and compare their p-values to an alpha of 5%, and if the $p - value < \alpha$, the country intercept is significant with a 95% certainty.

4.1.3 OLS diagnostic test - Robust Standard Errors

When performing the regressions in this section, we test for the OLS assumptions of heteroskedasticity and autocorrelation to get robust standard errors and valid hypothesis testing through our analysis. We use White's test in order to detect if heteroskedasticity is present. If only heteroskedasticity is present, we change into heteroskedasticity consistent standard errors. While in order to detect autocorrelation, we use Breusch-Godfrey's test. If we find evidence of both heteroskedasticity and autocorrelation, we change into heteroskedasticity and autocorrelation consistent standard errors.

4.2 Momentum Quantile Model - Looking for Systematic Relationships

4.2.1 Model Methodology

The momentum quantile model assigns rankings to the changes in ESG scores from the previous period (t-1) on a scale of 0-100. These ranks are then segmented into five quantile portfolios. Q1 contains companies with the highest positive change from the previous period, while Q5 includes those with the lowest or most negative change. Each portfolio are held until the next period, at which point both the returns and score change for each portfolio is analyzed. The model differentiates between two frequencies: quarterly, and annually. It also offers the flexibility to focus on specific ESG pillars.

4.2.2 How the model work to find systematic relationships

The model works by calculating and graphing the cumulative changes in ESG scores for each of the five portfolios, ideally displaying either a descending (5-1) or ascending (1-5) order for the cumulative change of each portfolio. If this pattern is observed, it signifies a systematic relationship between the portfolios. To ensure temporal consistency, we scrutinize the performance of each period using a bar plot like the one in Figure 1. It is worth noting that the model isn't designed to consistently produce a perfect 1-2-3-4-5 pattern. More typically, the most extreme portfolios (Q1 and Q5) exhibit the largest spread, while the remaining portfolios often show similar values and blend together. So, we will weigh Q1 and Q5 more heavily assessing whether we can call it systematic. Examples of the strong systematic tendencies we are looking for in Figure 1 and Figure 2.

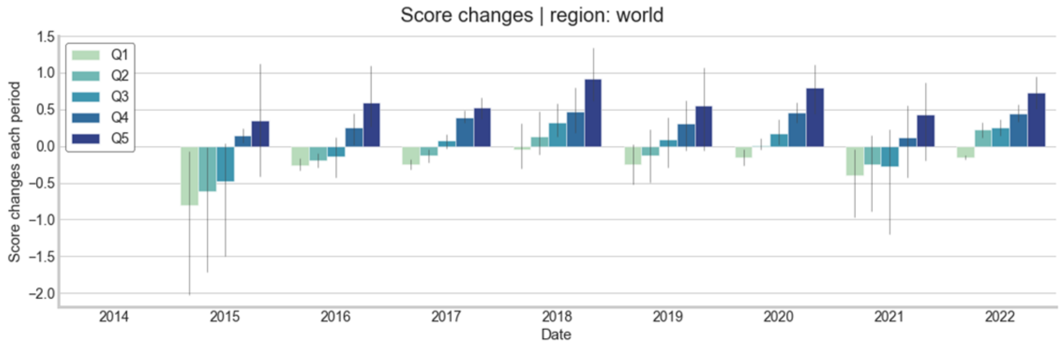

Figure 1: Example of strong systematic score changes

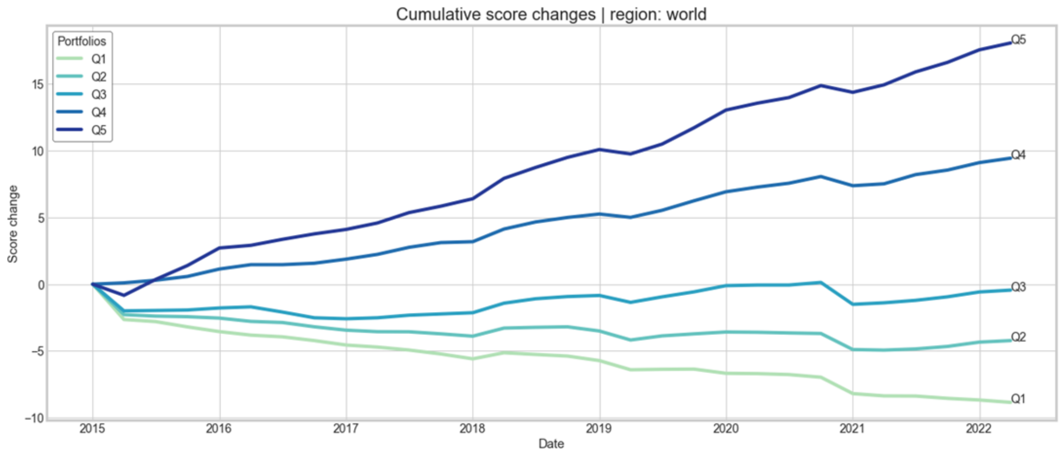

Figure 2: Example of strong systematic cumulative score changes

4.2.3 Clarify the model selection rationale and anticipated outcomes

The reason behind opting for this particular model is to delve into the dynamics of ESG score fluctuations. This method aids us in pinpointing systematic trends and ensuring consistency in our negative autocorrelation theory. Our intention is to use this model to verify our findings from the regression analysis and subsequently test it in a real-world setting. Observing a 1-5 pattern signifies that an increase in score is successively followed by another upward adjustment in the subsequent period. Conversely, a 5-1 pattern implies an initial uptick in the score is then followed by a downward adjustment in the next period, and vice versa. Also, here we apply the $n\lambda_i$ i.e., use change over longer periods and not only from the last period. The frequency (quarterly or yearly) determines how often to rebalance.

4.2.4 Hypothesis: Our expectations prior to implementing the model

Drawing from our regression analysis, we anticipate a (5-1) pattern. The Q1 portfolio, comprised of the companies that demonstrated the highest positive score change in the previous period, would undergo the most negative score shifts. On the other hand, Q5 should witness the most positive score alterations. These expectations stem from discovering significant negative coefficients in our regression analysis. In essence, an upswing is typically succeeded by a downward trend and vice versa.

4.3 Momentum Strategy Model - Can ESG momentum create excess returns?

4.3.1 Model Methodology

Our model is a refinement of the model outlined in Sankar et al. (2019), with additional adjustments and a more thorough analysis of the results. We construct three portfolios and begin by selecting the top 40% of the highest ESG scores within each industry. The rationale for our decision to use 40% rather than the 30% utilized by Sankar et al. (2019) stems from our interest in examining the impact of the negative momentum portfolio. This is also supported by Shanaev & Ghimire (2022), who found that the effects of finding abnormal returns are stronger in firms already considered ESG leaders when considering ESG momentum. Ouaknine et al. overlook this aspect, finding it necessary to merge the negative and neutral categories due to the scarcity of companies experiencing a 10% or greater decline each period. However, our data set encompasses more companies, and by incorporating additional 10% more companies, we can create a well-diversified negative momentum portfolio.

4.3.2 We define three portfolios

Positive momentum: Encompasses companies with a change in ESG score from the previous period of 10% or more. **Negative momentum:** Comprises companies with a negative change in ESG score from the previous period of 10% or more. **Neutral momentum:** Includes the remaining 40% that did not classify as positive or negative.

The model operates similarly to the Momentum Quantile Model, beginning with a lag of the ESG score (t-1). It then computes the percentage change on this lag from the previous period $\left(\frac{(t-1)-(t-2)}{t-2}\right)$. Instead of segmenting based on quantiles, it divides based on two criteria: a percentage change in score from the last period of at least 10% and a minimum required ESG score percentile within each industry. Only the top 40% of the highest scores within each industry are considered.

Table 1: Parameters - Momentum Strategy Model

Pillar	ESG, E, S, G and ESG ind adjusted
Frequency	Yearly and quarterly
Top percentile	Only use the top x % highest scores
Min. required change	Min required % score change
Window	How many periods change to look at
Data	Refinitiv world, MSCI EU, US, and world

The table shows the definitions of the different parameters we use in the momentum strategy model.

4.3.3 Clarify the model selection rationale and anticipated outcomes

This model has the potential to provide further evidence of momentum in ESG scores by examining whether the creation of momentum portfolios leads to generation of alpha returns.

4.3.4 Hypothesis: Our expectations prior to implementing the model

Based on prior research, we anticipate the positive momentum portfolio to outperform the others and the negative momentum portfolio to underperform. The justifications for these outcomes are often merely attributed to "positive momentum" and lack in-depth analysis. We intend to delve deeper. As per our regression analysis, a positive score change is typically followed by a downward scoring, implying that our positive portfolio should experience the most negative ESG changes during the holding period. Conversely, the negative momentum portfolio should see the most positive changes.

4.3.5 Factor Analysis

To evaluate the performance of our model and assess its ability to generate alpha, we employ a 3- and 5-factor model based on Fama & French (2014). We have chosen; WORLD/EU/US MSCI ESG leaders indices as benchmarks. These three indices will figure as our market portfolio in the factor models. Subtracting the risk-free rate to arrive at the market excess returns, we use the 3-month T-bill (US and German) divided by 12 as our risk-free rate in monthly terms.

The factors utilized in our study align with those proposed by Fama & French (2014). The size factor SMB, are constructed by ranking firms on size (market cap) from small to large. Average return of the three small-cap portfolios minus average return of the three large-cap portfolios, compose the factor returns. The value factor HML, are constructed by ranking firms on Book-to-Market (B/M) ratio from high to low. Composing the factor returns by subtracting average return of the two low B/M portfolios from average return of the two high B/M portfolios. While the profitability factor RMW, are ranked from robust to weak operating profitability (OP). Subtracting average return of the two weak OP portfolios from average return of the two robust OP portfolios.

Finally, the investment factor CMA, rank firms from conservative to aggressive investment style. Forming the factor returns from average return of the two conservative portfolios minus average return of the two aggressive portfolios. Importantly, these factors represent long-short factor returns, eliminating the need to adjust for the risk-free rate (Fama & French, 2014).

3-Factor Model:

$$R_{it} - R_{Ft} = \alpha + \beta_1(MKT_t - R_{Ft}) + \beta_2 SMB_t + \beta_3 HML_t + e_{it} \tag{5}$$

5-Factor Model:

$$\begin{aligned} R_{it} - R_{Ft} = \alpha + \beta_1(MKT_t - R_{Ft}) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t \\ + \beta_5 CMA_t + e_{it} \end{aligned} \tag{6}$$

Furthermore, we test if the intercept and the coefficients are significantly different from zero at the 5% significance level. This enables us to assess whether the strategy yields significant alpha after accounting for risk factors. Additionally, the model enables us to analyze the portfolio's factor exposures. Due to the fact that we use an equal-weighting scheme to compose the momentum strategy portfolios, we will expect to see positive exposure towards the size factor, meaning high exposure to small stocks, as we compare the strategy to a value-weighted index.

4.4 Momentum Trend Model – Can ESG momentum predict the next score adjustment

4.4.1 Model Methodology

Table 2: Parameters - Momentum Trend Model

Parameter	**Description**
Maximum error, MSE	Maximum means squared error in model, given by: $MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$
Minimum angle Ω	Minimum required steepness/angle of regression line, more than the output of function: np.rad2deg(np.arctan2(y, x))
Minimum angle long-term, ω	Minimum required steepness/angle of long-term regression line, $<$ np.rad2deg(np.arctan2(y, x))
Number of confirmations, Γ	Number of consecutive periods where all requirements must be met to generate a trend signal
Window of rolling regression, Δ	Number of periods used in the rolling regression
Window of rolling long-term regression, δ	Number of periods used in the long-term regression

The table shows the definition of the different parameters used in the momentum trend model.

We define an in-sample period (2013-2018) and one out-of-sample period (2019-2023). This allocation is a strategic move to minimize overfitting bias. The two data sets overlap from 2016-2019 because the model needs 3.5 years of data before it can start identifying trends. This means that the out-of-sample data starts in 2016, but the model does not start guessing trends before 2019, so there is no overlap where the model guesses. We needed to minimize the number of companies due to computer memory restrictions, so we will only look at US companies with ESG data for 2013 to 2022. We use the US because they have the highest coverage over the entire period; 1371 companies fulfilled these requirements.

4.4.2 Initiating the Financial Model with Rolling Regression

For each company, we start by doing a rolling regression over Δ number of periods. The dependent variable in this regression is ESG scores, while time is the independent variable. Following this, we compute the angle Ω of the regression line in the two-dimensional plane. By evaluating the regression line's angle, we gain insight into the trends. A positive angle implies an upward trend, neutrality around 0, and a downward trend if the angle is negative.

4.4.3 Incorporating a Higher Timeframe Trendline

Employing shorter period regressions, we can identify smaller trends within larger timeframes. In order to mitigate noise and the impact of the initial lag's negative coefficients observed in our regression analysis, we employ a longer period to offset the negative momentum effect. As a result, we incorporate an additional rolling regression on an extended timeframe, aiming to detect higher timeframe trends. This limits the model to guess a downtrend when we are in a bigger uptrend on a higher timeframe and vice versa. They both need to agree on the direction of the trend.

4.4.4 Trend Prediction Mechanism of the Model

Our model has six parameters that must be met before predicting an upward or downward trend. If any of these criteria are unmet, the model deems the available evidence insufficient for making a prediction and defaults not to guess, also called a neutral trend. The model utilizes past ESG scores exclusively to predict scores for the upcoming period, offering upward, downward, or neutral predictions.

4.4.5 Model training

Given that the model is formulated with six distinct parameters, each capable of assuming multiple values, it has been designed to be trainable. The model tests a variety of combinations for each parameter, documenting the statistics of each model iteration. Each training model is evaluated based on multiple factors, such as the accuracy of the predictions and the frequency of predictions while saving the values of each parameter. See Table 14 in the appendix for the last training set.

4.4.6 Clarify the model selection rationale and anticipated outcomes

Our objective is to ascertain whether ESG momentum can forecast an upward or downward adjustment in the subsequent period. If this can be achieved, it would support the existence of momentum in ESG scores. We have integrated other basic models as benchmarks as part of our approach. One model always predicts upward, another always downward, and the last model predicts randomly. None of these models incurs penalties if the score remains static, with penalties of -1 for incorrect predictions, 0 for neutral outcomes, and +1 for correct forecasts. We add these benchmarks because we want to compare the performance of our model against some simple models to assess our ability to provide meaningful predictions.

4.4.7 Hypothesis: Our expectations prior to implementing the model

Our hypothesis operates on the assumption of a single-period ESG change, predicting a positive shift following a negative one and vice versa. As per our understanding from the regression analysis and the momentum ranking strategy, this could potentially form an effective and simplistic model. However, our goal extends beyond solely predicting the direction of single-period changes.

Merely focusing on such predictions fails to provide insights into the overall trend of the company and where its score is headed in the future. Further, we have noticed that coefficients tend to lean towards the positive side over extended periods, suggesting that utilizing a broader data window might be beneficial.

4.4.8 Model Visualization

It can be hard to understand how all the components interact when working with complex models such as this trend-identifying model. This is why we have used visualizations throughout our research and models. The colored areas in the visualization in Figure 3 represent the model's predictions: green indicates an anticipated uptrend, red signifies a predicted downtrend, and yellow denotes neutrality, suggesting that the model lacks sufficient evidence to determine a trend conclusively. Purple lines denote the short-term rolling regressions, while the black line represents the ESG score. The blue dots signify the angle of the regression lines, whereas the grey line reflects the angle of long-term regressions. The red dots represent the error term for the short-term regression (not shown in the figure). We also employ a star-based rating system to assess the model's accuracy. A star value of '1' indicates a correct prediction, '0' signifies an ambiguous result, implying that the prediction was not outright incorrect but not precisely accurate either or no guess was made, and '-1' signifies an incorrect prediction.

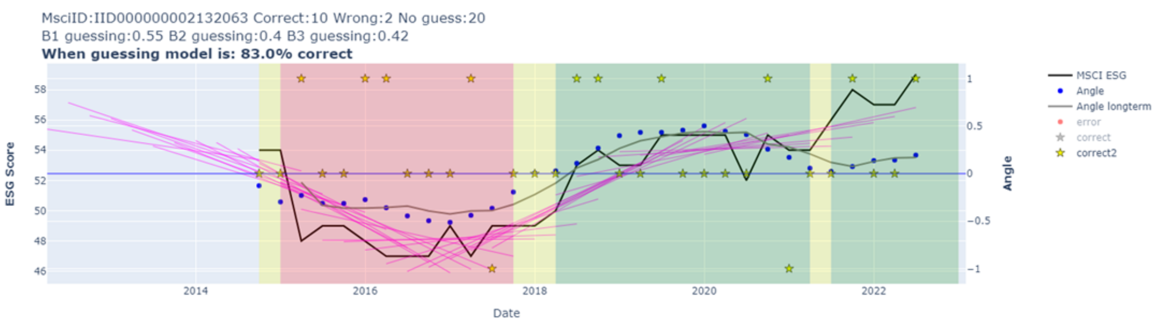

Figure 3: Model visualization

4.4.8.1 Understanding the High/Low Scoring System

First, observe the initial two '-1' star ratings during 2020-2021 in Figure 4, which were attributed due to the model's prediction of an uptrend while the score dropped for 2 consecutive periods. Then take note of the various downticks during the overall uptrend, which do not receive "-1" stars but are instead marked as neutral "0". This is because if a new high in the ESG score gets registered within the prevailing trend, these downswings get transformed into a "0". Due to the overall up trend still being intact. Had the uptrend not been interrupted in 2021 (signified by the yellow area), those two "-1" stars would have transformed into zeroes upon the record of the new high at the end of 2021. In the context of a downtrend, the arrival of a new low effectively neutralizes minor upswings within that trend.

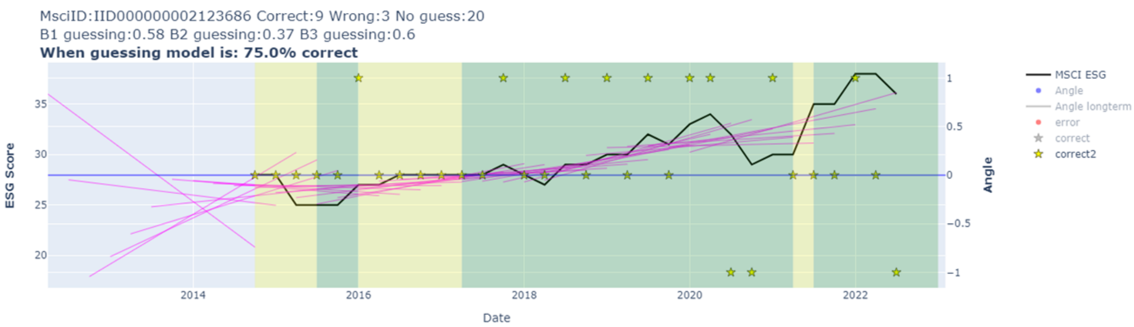

Figure 4: Example of how the High/Low scoring system works

A temporary but significant drop like the one we observe in the 2020-2021 period disrupts the preceding uptrend, which results in the recording of two wrong guesses at the end. However, if this fall in score had been just a little less the new high would have canceled out the two "-1" stars with zeros. This illustrates the complexity of achieving a high score. Without the brief interruption in the uptrend, the model's accuracy would have risen from 75% to 89%. To address this challenge, we could increase the number of data points in each regression, thereby fortifying the model against temporary trend disruptions. However, this enhancement does come with trade-offs. The model will become slower in adapting to new trends and exiting from existing trends.

Additionally, it needs a greater volume of data before it can commence making predictions.

4.4.9 Example of a typical 100%

The figure shows a typical good score company for the trend model. They are often characterized by a singular big trend, either up or down.

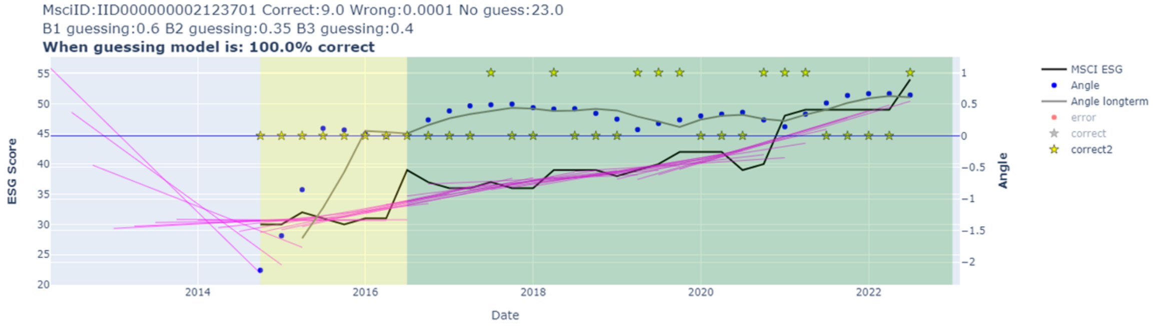

Figure 5: Example of a typical 100%

4.4.10 Analyzing two typical low-score instances

The first instance Figure 6 involves a company with a fairly flat score, making accurate predictions challenging for the model. This issue could be fixed by increasing the required angle for the regression, meaning the regression line must be at a steeper level for it to guess a trend. For this example, it would have resulted in no guesses and the whole plot would be in yellow.

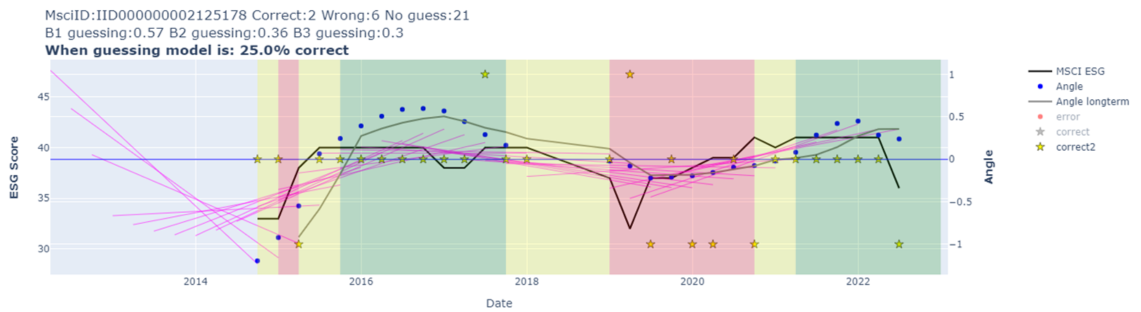

Figure 6: First example of a low score

The second case Figure 7 features a company that exhibits a mean-reverting or random pattern on a higher timeframe. This could be fixed by decreasing the number of data points in the regression, which makes the model more sensitive to upward and downward swings. In this scenario, it would result in steeper regression lines, and consequently, the model would make more predictions. However, as noted earlier, these adjustments bear implications on other preferred characteristics of the model.

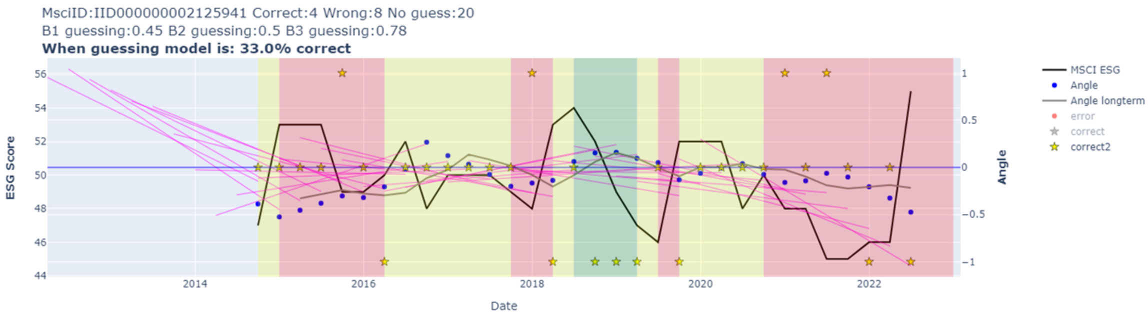

Figure 7: Second example of a typical low score

5 Data

5.1 Gathering of data

We primarily use MSCI's ESG scores as they are a leading provider of ESG data with over 40 years of experience measuring and modeling ESG performance. MSCI is recognized as a "Gold Standard data provider" and is best in class on SRI and governance research. They have the only data set with a live history of more than 13 years of history (*ESG Investing: ESG Ratings*, n.d.). The coverage is around 16 700 companies in 2022, 4576 in 2012, 1713 in 2007, and 420 in 1999. The companies come from various countries and can be differentiated by industry.

5.2 Other sources

When working with ESG scores it may be beneficial to use scores from different providers for several reasons. First, as it doesn't exist a fully standardized methodology on how to score companies, different providers may use different methods and criteria in their valuations which can lead to different scores for the same company (*GSIA |*, n.d.). Second, to reduce the risk of bias or subjectivity in the scoring process. Third, providers may choose to focus and overweight what they subjectively believe is important, using multiple scoring providers will give a more well-rounded view of the overall market view. Among other things, problems like these make scores from different providers disagree substantially. A paper by Berg et al. (2019) found that the correlations between six different raters are on average only 0.54. This incentivises us to consider other sources of data. As ESG data is relatively new and valuable, it is also hard to obtain without high costs According to Berg et al. (2019), MSCI and Refinitiv are the ESG data providers with the lowest correlation among all available providers, of 0.38. Hence, we will use these two scoring approaches in this paper to compare our hypothesis results and see how they diverge.

For historical stock prices, we will use EOD historical data (Support, 2021). This is a provider that also include stock prices for 14.000 delisted companies (*Landing Delisted — EOD*, n.d.). As the coverage is substantial, there is little to no survivorship bias. Referring to the tendency to consider the performance of current stocks or funds in the market as a complete representative sample while neglecting those that have failed or been discontinued. A bias that could lead to an overestimation of historical performance and general characteristics of a fund or market index (*What Is Survivorship Bias? Definition and Use in Investing*, n.d.).

We use the factor returns available from Kenneth French's website (*Kenneth R. French - Data Library*, n.d.) with a 3- and 5-factor model. We will be using US factor returns from; small-minus-big (SMB), high-minus-low (HML), robust-minus-weak (RMW) and conservative-minus-aggressive (CMA) (Fama & French, 2014). In order to select an appropriate benchmark index for our analysis we substitute the Kenneth French market index (MKT-RF) with three different indices for three distinct cases, World – MSCI World ESG Leaders, U.S. – MSCI US ESG Leaders, and Europe – MSCI Europe ESG Leaders (*MSCI ESG Leaders Indexes*, n.d.).

Our use of MSCI ESG scores influences our choice of indices. Since we limit the investment universe to ESG leaders, we want a benchmark that considers the same ESG leaders for the most appropriate comparison. Due to the width dispersion in ESG scores, it makes the most sense to use MSCI's indices as they are based on identical scores. Since no equal-weighted indices where available, we are limited to value-weighted indices. Furthermore, we calculate the excess returns for benchmark indices by subtracting the U.S. 3-month T-Bill for both the world and U.S. indices. We use the U.S. T-Bill for the world index as U.S. bond market (ticker=IR3TIB01USM156N) (Organization for Economic Co-operation and Development, 1964) is the best proxy for the risk-free rate internationally as the country preserves the world reserve currency. Hence, the closest to the risk-free asset (Damodaran, 2008). Germany is one of few countries that is rated AAA by S&P, Moody's and Fitch (Zoppe & Lenzi, n.d.), and additionally has the most liquid bonds in Europe (European Central Bank., 2022). Therefore, we use the German 3-month T-Bill (ticker=IR3TIB01DEM156N) (Organization for Economic Co-operation and Development, 1960) as our risk-free rate for the Europe index.

5.3 Handling data

We have noticed some gaps in the timeline of the ESG data. This poses challenges since our approach relies on changes over time. To manage this, we have iterated through each company based on dates, and when a gap is spotted, we assign a new ID to the company to distinguish them. See Figure 33 in the appendix for the code snippet.

5.4 Data period

Even though the MSCI data set has a monthly frequency, scores are not updated as often. We will, therefore only look at quarterly and yearly data. The Refinitiv data set reports yearly. Hence we will use yearly frequency for that data set. The fact that scores from before 2013 are backfilled reduces the quality of the data drastically. Further, we see that the focus on the ratings increases with time alongside the quality of the data. As we want a long period of data, we must keep in mind that we only start with 2.5% (1999) of the companies we have available in 2022 for the MSCI data set. In 2013 the MSCI data set experienced a significant increase in companies available, jumping from 4659 to 10,842 companies and 15,667 companies in 2022 (Figure 18). For the Refinitiv data set, the increase in companies is more drastic toward the end but provided with 3335 companies in 2013 and 11.366 companies in 2022

(Figure 19). Based on these facts, we decided to work with a sample period from 2013 onwards. There are also examples of other studies using the same starting period. The study of Bekaert et al. (2023) used a sample period from 2013 to 2018.

5.5 Data restriction in models

Given the extensive nature of our data, coupled with numerous variables such as frequency, ESG pillars, geographical region, and factors at the company, industry, and country level, it presents an overwhelming number of combinations to examine. In the analysis of the three main models, we will focus on the EU, US, and global scales across all levels (company, industry, and country), while narrowing the frequencies and pillars to those only deemed significant from our regression analysis.

5.6 Construction and Definitions of scores

5.6.1 MSCI Scores

The MSCI ESG score is divided into three pillars; environmental, social, and governance which are further divided into ten themes, see Table 3. The environmental pillar has four themes: climate change, natural capital, pollution & waste, and environmental opportunities. The social pillar is divided into; human capital, product liability, stakeholder opposition, and social opportunities. While the governance pillar has two themes; corporate governance and corporate behavior. These themes are furthermore divided into a total of 35 ESG key issues. In order to arrive at the final ESG score, they use the weighted average of the environmental and social key issues scores. The governance pillar is computed separately and has a fixed weight for all industries, floored at 33% as a minimum. "MSCI ESG Ratings Methodology" (2022).

Table 3: MSCI ESG Ratings Methodology, 2022

3 Pillars	10 Themes	35 ESG Key issues	
Environment	Climate Change	Carbon Emissions	Fin. Environmental Impact
		Product carbon Footprint	Climate Change vulnerability
	Natural capital	Water Stress	Raw Material Sourcing
		Biodiversity & Land Use	
	Pollution & Waste	Toxic Emissions	Electronic Watse
		Packaging Material & waste	
	Environmental Opt.	Opt in Clean tech	Opportunities in renewable
		Opt in Green Building	Energy
Social	Human Capital	Labor Management	Human Capital Dev.
		Health & Safety	Supply chain labour
	Product Liability	Product Safety & Quality	Privacy & Data Security
		Chemical Safety	Responsible Investment
		Consumer Fin. Protection	Health & Demographic Risk
	Stakeholder Opposition	Controversial Sourcing	
		Community Relations	
	Social Opportunities	Access to Communications	Access to Health Care
		Access to Finance	Opt. in Nutrition & Health
Governance	Corporate Governance	Ownership & Control	Pay
		Board	Accounting
	Corporate Behavior	Business Ethics	
		Tax Transparency	

5.6.2 Refinitiv Scores

As for MSCI ESG scores, the Refinitiv ESG scores are divided into three pillars; environmental, social, and governance, aggregated by ten category scores. Emission, innovation, resource use for the environmental pillar, human rights, product responsibility, workforce and community for social pillar, and for the governance pillar; management, shareholders and CSR strategy. These are again aggregated on between 70 and 170 booleans (yes/no) data points relevant for each specific industry, based on only disclosed and publicly available data. To formulate a final ESG score, they use their own Refinitiv ESG Materiality Matrix. A matrix calculates category weights for the environmental and social pillars based on an objective and data-driven approach to determine each theme's relative importance towards each industry group. They use two methods to calculate the matrix: an industry median method for numeric data or transparency weights for boolean data. Based on the importance of each industry, the category weight is distributed from 1-10. While the governance pillar has a default weight of 15. Due to the use of the industry median method, the scores become industry adjusted. Addition- ally, the fact that their model is fully automated, data-driven, and transparent, makes it immune to bias towards subjectivity and hidden calculations or inputs ("Environmental, Social and Governance (ESG) Scores from Refinitiv", 2020).

5.7 Descriptive Statistics

Below in Table 4, some standard descriptive statistics for the two data sets. Comparing the two parts of the table, we observe that scores are on average slightly higher for the MSCI scores than for Refinitiv scores. Meanwhile, the dispersion is higher for the Refinitiv scores compared to the MSCI scores. This may come as a result that the two providers use different methodologies for calculating the scores or weighing scores differently.

Table 4: MSCI & Refinitiv - Descriptive Statistics

Data set	ESG	E	S	G	
MSCI	count	162648	162648	162648	162648
	mean	46.876	49.795	45.408	51.336
	std	12.154	22.322	16.765	19.525
	min	0.000	0.000	0.000	0.000
	25%	39.000	34.000	35.000	39.000
	50%	47.000	49.000	46.000	50.000
	75%	54.000	65.000	56.000	64.000
	max	98.000	100.000	100.000	100.000
Refinitiv	count	80857	80857	80857	80857
	mean	42.722	34.593	43.103	49.258
	std	20.692	28.750	23.693	22.480
	min	0.402	0.000	0.053	0.055
	25%	26.080	6.465	24.093	31.192
	50%	40.884	30.055	40.846	49.602
	75%	58.497	58.587	61.160	67.388
	max	95.448	99.223	99.564	99.622

The figure shows descriptive statistics for MSCI & Refinitiv ESG scores.

From the pie-charts in Figure 20 and Figure 21 we see how the data is distributed over industries for the MSCI and Refinitiv data sets respectively. In total, there are 89 unique industries in the data set for MSCI. But we know there are some overlaps between equal industries in different regions with slightly different names, increasing the total number of industries. Looking at Figure 20 we see that the banking industry is the dominant industry by a good margin. Behind, we find utilities, Software & Services and Real Estate Management & Services. Meanwhile, for the Refinitiv data set, shown in Figure 21,there are 69 unique industries. As in the MSCI data set, the banking industry is the largest but substantially less dominant. Metals & mining is the second largest industry, and Biotechnology is the third largest. As in the MSCI data set, the Real Estate Management & Development is one of the largest industries, ranking fourth in the Refinitiv set. The Refinitiv data set is more evenly distributed across industries.

Moving over to the countries, the pie-chart in Figure 22 shows a total of 138 unique countries in the MSCI data set. The US is the dominant country as it covers for more than $\frac{1}{4}$ of the companies. Japan (JP), Great Britain (GB) and China (CN) following behind. While Figure 23 shows that the Refinitiv data set provides us with 93 countries in total, where we see the same dominance from the US with China, India, Great Britain and Japan behind.

6 Results and analysis

6.1 Regression Analysis

6.1.1 Overall findings from regression analysis

Our regression analysis examining the MSCI and Refinitiv data sets reveals a significant and mostly negative impact from λ's on change in ESG scores (Table 8). Showing an inverse relationship and thus implying a reversing in scores rather than momentum. FFurthermore, a consistent pattern exists throughout the pillars that $4\lambda_Q$ and $1\lambda_Y$ is the most impactful λ's. However, ESG $1\lambda_Y$ and governance $1\lambda_Y$ show evidence of the largest coefficients, -0.22 and -0.19, respectively. Meaning, when a score increases 10%, it will on average decrease by 2% in the next period. Summarizing these results, we find that yearly frequency concerning the ESG and governance pillar scores might be the appropriate testing objective in the forthcoming. However, in a broader picture, the general size of the coefficients is relatively low, which ultimately forces us to be careful about stating the magnitude of the impact. Hence, we should expect weaker trends due to these relatively low coefficients.

Delving deeper into the MSCI and Refinitiv data sets, it becomes apparent that there are distinct industry effects on a company's ESG score changes across quarterly and yearly frequencies. Looking at Table 7, the MSCI data uncovers significant industry effects across both frequencies. However, the number of significant industry intercepts is exceptionally higher for quarterly frequency (Table 10) than for yearly frequency (Table 12), which displays only a few significant industries in the case of overall ESG score and environmental pillar score. Similarly, in Table 12, the Refinitiv data provides evidence of significant industry effects but is less revealing for in-depth exploration due to a lack of significant industry intercepts and consistent patterns. Hence, quarterly frequency is a more appropriate frequency to investigate the industry effect with our models.

The data sets tell another story when considering the geographic lens. Both MSCI yearly frequency and the Refinitiv data sets reveal significant geographical effects on a company's ESG score changes. A crucial distinction emerges in the MSCI data set; specific national environments significantly influence score changes. Showing a prevalence of positive country intercepts for overall ESG score, environmental and social pillar scores. This could, for example, be from a generally beneficial effect of country policies, such as tightening environmental regulations. On the other hand, the governance pillar presents a contrasting view, indicating a negative country effect. This could be a result of strict policies on governance in those countries. While in the Refinitiv data set, no single country stands out to have a significant impact.

We note that the Refinitiv scores provide us with limited information for further exploration of the scores concerning our research question. While some observations can be made about industry effects, the lack of significant industries and countries and weak significant and small coefficients make it challenging to draw concrete conclusions or develop comprehensive strategies.

6.1.2 Pooled Regression

6.1.2.1 Quarterly frequency – MSCI data

From the results we obtained from the pooled regression in Table 7, we reject the null hypothesis that all λ coefficients are equal to zero for all three pillar scores ($F\text{-statistic}_E = 19.94, F\text{-statistic}_S = 21.63, F\text{-statistic}_G = 64.95$) as well as for the overall ESG score ($F\text{-statistic}_{ESG} = 80.86$) for quarterly frequency at the 5% significance level. Indicating that at least one of the λ's has a significant impact on change in scores, indicating that previous changes in scores impact future ones.

One notable finding (Table 8) is that the overall ESG and social and governance pillar scores display negative coefficients. This is quite interesting as it tells us that an increase in the λ results in a decrease in the change in score, and vice versa. Hence, there appears to be an inverse relationship between λ's and change in score. Economically, this could be explained by companies decreasing their focus on the theme as they increase their score, hence getting punished after an increase. Further, looking at Table 8, we find that $4\lambda_Q$ is significant for all three pillar scores and the overall score. In addition to be the largest coefficient for all three pillar scores $4\lambda_Q$ (E=-0.087, S=-0.0105, G=-0.0125). However, the ESG score shows that $3\lambda_Q$ (-0.0186) is the largest coefficient. We can interpret the result as an indication that $1\lambda_Y$ has the most substantial impact on changes in score especially in the case of the pillars, and therefore also the most appropriate frequency to use for our models on company-specific momentum.

Additionally, we do also find the intercept significant and positive for all three pillars as well as the overall ESG score. Implying that even without the effect of the λ's scores will have a baseline positive change, meaning that on average scores increase over time extracting the effects of λ's. The economic interpretation of this may be that over time, scores increase due to companies increasing focus on the theme.

6.1.2.2 Yearly frequency – MSCI data

Moving on to yearly frequency, we see the same results (see: Table 7) as for quarterly frequency. We reject the null hypothesis for all three pillar scores (F-statistic$_E = 2.97, F$-statistic$_S = 5.08, F$-statistic$_G = 8.26$) as well as for the overall ESG score (F-statistic$_{ESG} = 40.00$) at the 5% significance level. Meaning that we have at least one λ in ESG and all three pillars that has a significant impact on change in scores.

Examining the significance of λ's in Table 8, we find that $1\lambda_Y$ is significant for all scores except for the environmental (E) pillar score. Meanwhile, $7\lambda_Y$ showed significance for all scores except the governance (G) pillar score. The above results tell us that $1\lambda_Y$ and $7\lambda_Y$ are the most insightful explanatory variables in explaining changes. We do observe the same pattern as in quarterly frequency, that $1\lambda_Y$ ($4\lambda_Q$) stand out as the most impactful independent variable. Especially for ESG and governance $1\lambda_Y$, with coefficients of -0.22 and -0.20. Meaning, if the previous change were a 10% decrease, the score will increase by approximately 2% in the subsequent period. A closer look at the coefficients of the overall score reveals a positive relationship for $5\lambda_Y$ and $7\lambda_Y$, implying that by extending the period of change, we could achieve a slightly positive impact from λ on change. These results show that significant coefficients for yearly frequency tend to be larger than those for quarterly frequency. This is in line with the discussion from 6.1.2.1 that yearly frequency withholds more information in terms of explaining changes in scores. (Table 8).

6.1.2.3 Yearly frequency – Refinitiv data

Progressing to the Refinitiv scores, we reject the same null hypothesis as we did for quarterly and yearly frequency in the MSCI data set for the overall ESG score ($F\text{-statistic}_{ESG} = 1.89$) as well as for social and governance pillar scores ($F\text{-statistic}_S = 3.61, F\text{-statistic}_G = 4.60$) (Table 7). Stating that there is at least one λ has a significant impact on change in the score at the 5% significance level. However, the results obtained in Table 8show that these results differ from those of the MSCI scores, as we observe different λ's as significant. On the other hand, we are not able to reject the null hypothesis for the environmental pillar score ($F\text{-statistic}_E = 0.05$). This tells us that all tested λ's are useless in explaining the variation of change in the environmental pillar score. Looking at the ESG, social $2\lambda_Y$, and governance $4\lambda_Y$, these displayed positive coefficients. Suggesting that the most distant significant λ_Y , contributes positively to the change in score. Indicating that if we extend the length of λ's there will be opportunities for positive relationships with the change in score. And thus, potentially find positive momentum in scores over longer periods. Likewise, as for the MSCI scores, we observe a significant and positive intercept for all three pillar scores and the overall score (Table 8).

6.1.3 Industry-Fixed Regression

6.1.3.1 Quarterly frequency – MSCI data

We observe that the overall ESG score and all three pillar scores exhibit significant industry effects. Hence, rejecting the null hypothesis (F-statistic$_{ESG}$ = 23.71, F-statistic$_E$ = 12.14, F-statistic$_S$ = 24.90, F-statistic $_G$ = 4.67, Table 9) that there are no industry-effects. This result is consistent with our hypothesis that industry drives change in the company scores. Looking at industry specificity in Table 10, the environmental pillar exhibits the highest number of significant industries. Suggesting that it is the most affected pillar, by industry-specific regulations. While the governance pillar is least affected by industries, as it has the least amount of significant industry intercepts. Furthermore, the overall ESG score and social pillar score both show an overweight of positive industry intercepts. Implying that industry has a mostly positive effect on firms' changes in social and overall ESG scores. Meanwhile, environmental and governance pillar scores demonstrate an overweight of negative industry intercepts, indicating a primarily negative impact from the industry effect.

6.1.3.2 Yearly frequency – MSCI data

We detect the same result for yearly frequency as we did for quarterly frequency in the MSCI data set. Rejecting the null hypothesis (Table 7) that there are no industry effects for the overall ESG score (F-statistic$_{ESG}$ = 157.21) as well as for all three pillar scores (F-statistic$_E$ = 34.05, F-statistic$_S$ = 75.26, F-statistic$_G$ = 36.83) at the 5% significance level. However, contrary to quarterly frequency, few industry intercepts are significant (see Table 12). Only the overall ESG score and environmental pillar score display those few significant industry effects.

6.1.3.3 Yearly frequency Refinitiv data

In our analysis of the Refinitiv data set, we observe significant industry effects (Table 7) across all three pillar scores (F-statistic$_E = 5.00, F$-statistic$_S = 15.27, F$-statistic$_G = 19.99$) as well as for the overall score (F-statistic$_{ESG} = 17.77$). Hence, we reject the null hypothesis that there are no country effects at the 5% significance level. Contrary to the MSCI data sets, we see from Table 12 that the Refinitiv data shows no evidence of significant industry intercepts. The absence of significant industry intercepts could be rationalized through the fact that the Refinitiv scores are industry-adjusted (chapter 5.6.2). Therefore, which industry the firm is placed within should be irrelevant to the score.

6.1.4 Country-Fixed Regression

6.1.4.1 Quarterly frequency – MSCI data

Due to pc memory restrictions, we could unfortunately not run the country-fixed regression.

6.1.4.2 Yearly frequency – MSCI data

Over to the country fixed effects, we observe from the regression that the overall score (F-statistic$_{ESG} = 22.57$)) and all three pillar scores (F-statistic$_E = 27.91, F$-statistic$_S = 19.26, F$-statistic$_G = 36.77$) exhibit significant country effects at the 5% significance level (Table 7). Hence rejecting the null hypothesis that there are no country-specific effects, demonstrating the importance of geographical location for a company's changes in the score. Several countries appear as significant intercepts in all scores, indicating that certain national environments have a more substantial influence on changes in scores. Looking at Table 13, the environmental pillar score especially stands out as the score with the highest number of significant country intercepts. An indication of strong influence of countries on this specific pillar.

In terms of the sign in front of country intercepts, Table 13 displays that the overall score along with environmental and social pillars, shows an overweight of positive country intercepts. This suggests that being located in these countries has a positive impact on changes in scores for companies. In contrast, the governance pillar presents an overweight of negative country intercepts, indicating that the governance scores of firms are adversely affected by being located in these specific countries. This could be explained by governmental policies within the different pillars and how strict the policies are in a company's specific country of operation.

6.1.4.3 Yearly frequency – Refinitiv data

As observed for MSCI scores, the same is observed for Refinitiv scores. We reject the null hypothesis that there are no country effects for the overall ESG score (F-statistic$_{ESG} = 18.76$) and the three pillar scores (F-statistic$_E = 5.25, F$-statistic$_S = 10.26, F$-statistic$_G = 17.32$), displayed in Table 7. Hence, there are significant country effects at the 5% level, underscoring the impact of geographical location on the changes. What is interesting is the fact that there are no significant country intercepts across any of these scores. An indication that no single country significantly impacts scores, even though there are weak country effects.

6.2 Momentum Quantile Model

6.2.1 Summary of Quantile Model results

This study on portfolios classified based on ESG score trends shows systematic patterns in ESG score changes but no consistent patterns in returns. Portfolios containing companies with the highest ESG score declines (Q5) exhibited the most positive adjustments during the holding period, in line with the negative coefficients found in regression analysis. Conversely, portfolios with companies exhibiting the most ESG growth (Q1) in the previous period experienced the greatest score decreases. We did not find the same systematic structures when we tried to do the same with industry portfolios. Nevertheless, we did see a systematic effect for the EU country-level portfolios. Meaning, if the average ESG score for a country gets downgraded, it will likely be adjusted upwards in the next period. The results point out that while there are systematic patterns in ESG scores, these do not translate into systematic patterns in returns, suggesting that ESG score changes alone may not be a reliable indicator of financial performance.

Table 5: Momentum Quantile Model - Summary

Level	**Score Change**	**Coefficient**	**Systematic?**	**Data**
Company	ESG $3\lambda_Q$	-0.0186	High, (Q5-Q1)	World, MSCI
Company	ESG $1\lambda_Y$	-0.2199	High, (Q5-Q1)	World, MSCI
Company	ESG$1\lambda_Y$	-0.0814	None	World, Refinitiv
Company	ESG $2\lambda_Y$	0.0527	None	World, Refinitiv
Company	ESG $5\lambda_Y$	0.0465	Low	World, MSCI
Company	S$1\lambda_Y$	-0.0364	None	World, MSCI
Company	G$1\lambda_Y$	-0.1999	Medium-High, (Q5-Q1)	World, MSCI
Industry	ESG$1\lambda_Y$		None	World, MSCI
Industry	ESG $1\lambda_Y$		None	EU, MSCI
Country	ESG $1\lambda_Y$		None	World, MSCI
Country	ESG $1\lambda_Y$		Medium-High, (Q5-Q1)	EU, MSCI

Overview of the results from the Mommentum Quantile Model.

6.2.2 Company level

These portfolios consist of stocks with equal weight.

6.2.2.1 Company | ESG3$\lambda_\mathbf{Q}$ | Coefficient : −0.0186 | World, MSCI

Upon examining the ESG evolution in the portfolios, we discerned a systematic pattern from Q5 to Q1 (Figure 8). The Q5 portfolio, comprising companies with the steepest decline in score over the past three quarters, exhibited the most upward adjustments during the holding period. This is in line with the negative coefficient. Conversely, we observed a reverse effect in the Q1 portfolio, suggesting that companies showing the most growth in the previous period suffered the greatest decreases in the next. We made an interesting discovery when the model was applied with an insignificant change like ESG $1\lambda_Q$ and ESG $2\lambda_Q$; it disrupted the perfect systematic pattern. This pattern only remains intact at the significant frequency we discovered through the regression analysis (at least for the quarterly frequency).

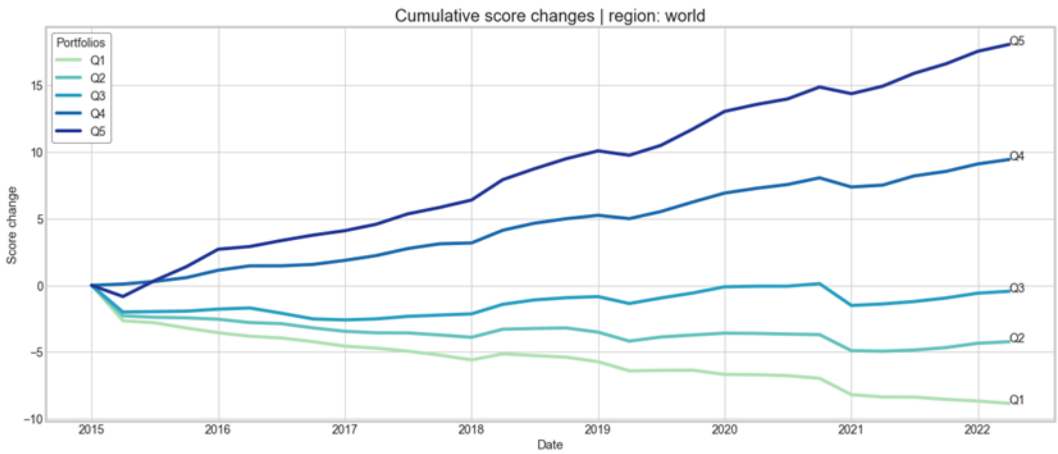

Figure 8: Cumulative ESG score change for $3\lambda_Q$

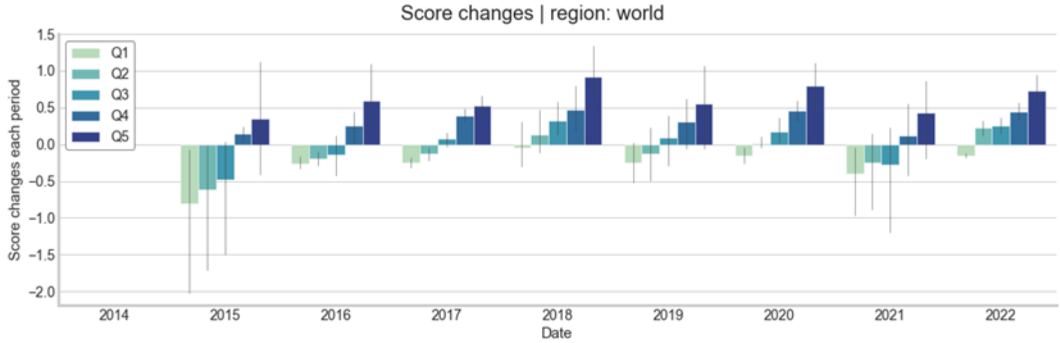

Figure 9: Barplot of ESG score change for $3\lambda_Q$, showing temporal consistency in systematic patterns

We do not find enough evidence in the returns to conclude any specific patterns or systematics (Figure 10). The lower quantile portfolios Q1 and Q2 underperform some, but it might be due to luck, and we will call it inclusive. This implies that negative score adjustments might be correlated to lower returns.

Figure 10: Cumulative returns for $3\lambda_Q$

Surprisingly, the Q1 and Q5 portfolios have relatively low turnover, around 50% each period (Figure 11). This means that half of the stocks continue their ESG change momentum. Which is contradicting to what we know about the negative coefficient from the regression analysis. Unless the new stocks that entered the portfolio move stronger in the opposite direction than those that continue their momentum.

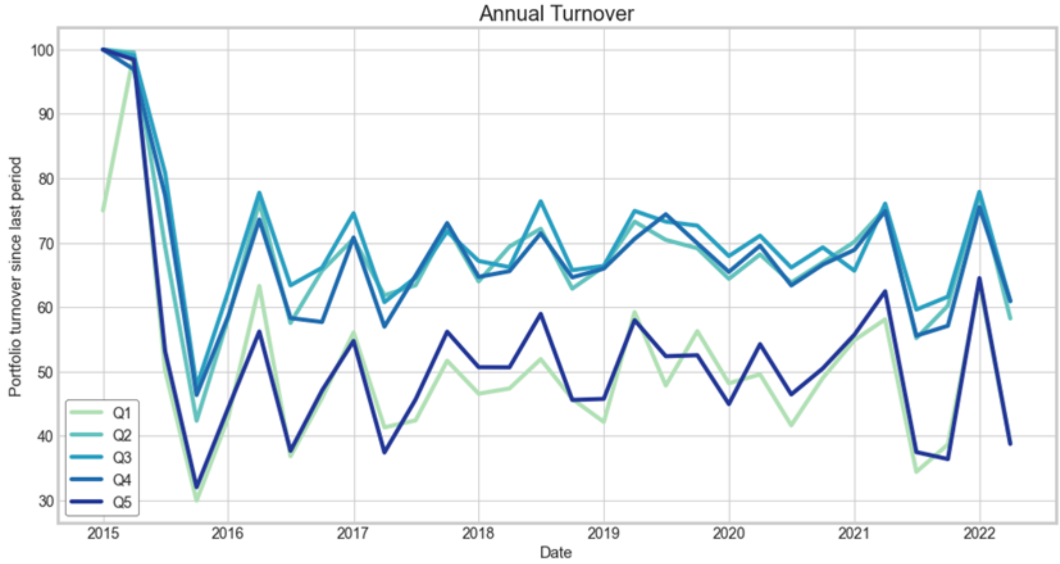

Figure 11: Turnover for $3\lambda_Q$

6.2.2.2 Company | $\mathbf{ESG1\lambda_Y}$ | Coefficient : -0.2199 | World, MSCI

Much of the same results from ESG $3\lambda_Q$. No systematic regarding returns, while we have strong systematic patterns in ESG score change.

6.2.2.3 Company $\mathbf{ESG5\lambda_Y}$ | Coefficient : 0.0465 | World, MSCI

Q5 continued to witness the most positive ESG score changes, ending at 6 in cumulative ESG change over the period. Interestingly, Q1 to Q4 also registered positive scores, with Q2 experiencing the smallest cumulative score just above zero. Except for the Refintiv data, this is the only time we have seen all portfolios ending their cumulative score above 0. While there were limited systematic trends observed, the negative changes in Q1 and Q2 vanished as the coefficients barely started turning positive. No systematic tendencies were found in returns.

6.2.2.4 Company | $\mathbf{G1\lambda_Y}$ | Coefficient : -0.1999 | World, MSCI

Here, the Q5 score increased by approximately 10, maintaining a systematic pattern. Q1 to Q3 scores remained close around 0 but still in systematic order. As for returns, no systematic pattern emerged as all portfolios ranged from 40 to 50% in total cumulative return over the period, with no specific order. This shows that the G pillar is in fact a big driver of ESG scores.

6.2.3 Industry-Level Assessment

Our industry portfolios are composed of equal-weighted industry portfolios, using ESG $1\lambda_Y$ to rank. Similar to previous model runs, Q5 saw the highest cumulative score change over the period. No systematic patterns were discerned for scores or returns. Each portfolio contained about 13 industries. This result was consistent across the world and Europe. The Refinitiv data showed no systematic effects, with all portfolios experiencing positive score changes on average, much like the company-level observations. Surprisingly, Q1 saw the highest increase during this period, which may have occurred randomly given the lack of systematic tendencies.

6.2.4 Country-Level Assessment

The country portfolios comprised of equal-weighted country portfolios, using ESG $1\lambda_Y$ to rank. No systematic patterns for scores or returns were found when applied globally. However, an almost perfect systematic relationship emerged when we limited the model to Europe: Q1 decreased by 6 scores, Q5 increased by 6, while Q2 to Q4 remained around 0. This suggests that European countries exhibit stronger indications of the mean reverting pattern. Each portfolio consisted of 5-6 countries at any given point. For the Refinitiv data, neither returns nor scores were systematic, with all portfolios experiencing an increase in scores.

6.3 Momentum Strategy

We will only test for $n\lambda_i$ with high systemic tendencies from the Quantile model as we believe them to be the best candidates for building momentum portfolios due to their predictable score pattern. To do an in-depth analysis, we have chosen one $n\lambda_i$ that best represents the group of tests in terms of similarity. For the quarterly frequency portfolios, we look at three quarters change; we have therefore lowered the required percent change from 10% to 7.5% since we use $\frac{3}{4}$ of a year.

6.3.1 Overall results

Our analysis of different ESG investment strategies across the World, EU, and US markets from 2015 to mid-2022 revealed several noteworthy trends. The World ESG Leaders Benchmark consistently surpassed momentum portfolios, with the negative momentum portfolio typically outperforming the positive one. The positive momentum portfolio, however, registered the only significant but negative alpha. Both portfolios exceeded the benchmark in the EU market, yet no significant alpha was recorded. The negative momentum portfolio generally performed well against the benchmark and the positive momentum portfolio. The same can be said for the US market, except that the negative portfolio was the only one to outperform the US benchmark without significant alpha.

A notable trend was that the ESG scoring patterns were consistent across all markets, with the positive portfolios typically experiencing score decreases while the negative portfolios seeing increases. The negative momentum portfolio also showed similar performance in scores to the portfolio containing the bottom 60% lowest ESG companies.

Overall, we found little evidence that building portfolios based on ESG momentum will yield excess returns. Despite fluctuations in individual portfolio performance, the negative momentum portfolio generally performed well across different regions, although its successes are not indicative of higher risk-adjusted returns. We need to point out that our benchmarks are value-weighted while the portfolios are equal-weighted which can make it harder to produce significant alphas.

Table 6: Momentum Strategy Model - Factor results

Score Change	**Coefficient**	**Data**	**Portfolio & Model**	**Alpha (monthly)**
ESG $3\lambda_Q$	-0.0186	World, MSCl	**Pos-3-F**	**-0.005411***
			Pos-5-F	-0.004347
			Neg-3-F	-0.003332
			Neg-5-F	-0.003200
		US, MSCI	Pos-3-F	-0.005169
			Pos-5-F	-0.003997
			Neg-3-F	-0.003781
			Neg-5-F	-0.004067
		EU, MSCl	Pos-3-F	0.000499
			Pos-5-F	0.001377
			Neg-3-F	0.004497
			Neg-5-F	0.004645
ESG $1\lambda_Y$	-0.22	World, MSCl	**Pos-3-F**	**-0.004153***
			Pos-5-F	-0.003863
			Neg-3-F	-0.001230
			Neg-5-F	-0.000728
		US, MSCl	Pos-3-F	-0.003665
			Pos-5-F	-0.003283
			Neg-3-F	0.004185
			Neg-5-F	0.005220
		EU, MSCl	Pos-3-F	0.003742
			Pos-5-F	0.003974
			Neg-3-F	-0.000215
			Neg-5-F	-0.000080
		World, Refinitiv	Pos-3-F	0.005354
			Pos-5-F	0.003127
			Neg-3-F	0.001375
			Neg-5-F	0.056021

Alpha with * is significant at 5%. Alpha is in monthly terms.

6.3.2 Factor Analysis

The results from the factor analysis shown in Table 6, display that the positive momentum portfolio shows a significant alpha at the 5% significance level but only in the case of the 3-factor model. In this case, it is negative, which economically means that the portfolio performs a negative excess return in relation to the benchmark index, SMB- and HML-factor. Looking at the other factor coefficients, there is a clear pattern that both negative and positive portfolios have a significant tilt toward small cap stocks. None of the other factor coefficients display any significant exposure on the portfolios. The high exposure towards the SMB-factor is expected since the momentum portfolios are equal-weight while the benchmark is value-weighted. This causes our portfolios to be biased towards smaller stocks.

6.3.2.1 ESG $1\lambda_{\mathbf{Y}}$ | BM : WORLD ESG LEADERS

Overall, the positive portfolio stood out for its significant alpha in the 3-factor model, scoring a monthly alpha of -0.004153 (Table 6). The negative portfolio outperformed all other portfolios except for the benchmark (Figure 12). This is not surprising as the World ESG Leaders is a value-weighted index with a high concentration of US stocks known to outperform the rest of the world. Since the portfolios are equally weighted it consists of more stocks from underperforming regions relative to the US. The momentum portfolios only outperformed in 2016 and 2022 (Figure 13).

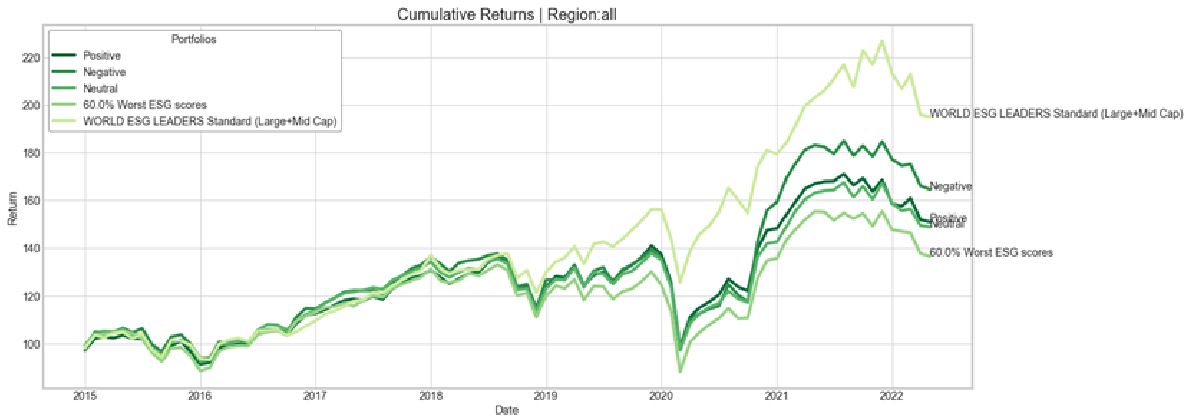

Figure 12: Cumulative returns for $1\lambda_Y$

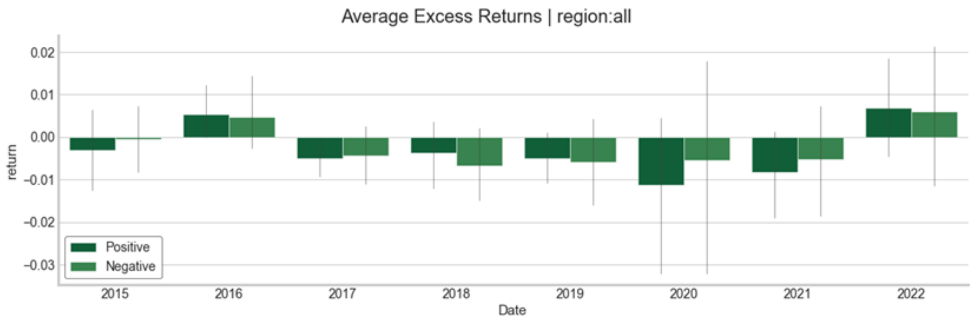

Figure 13: Excess returns over benchmark

The ESG score patterns remained consistent; the positive portfolio experienced mostly negative changes about (-9), largely due to REITS's, tobacco, Hungary, and Israel. While the negative portfolio underwent mostly positive adjustments and ended with (5) in cumulative score change due to the construction materials, Wireless telecom industries, Austria and the Czech Republic. See Figure 24 to Figure 32 for the full industry and country contributions to returns and scores.

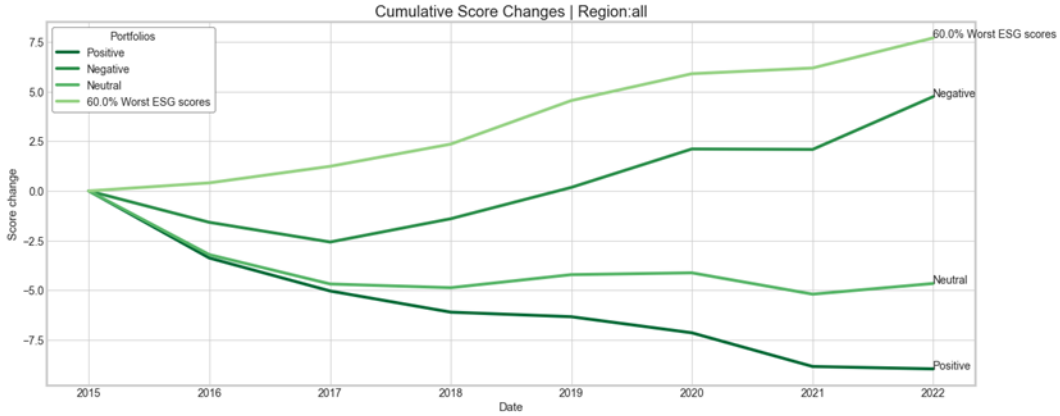

Figure 14: Cumulative score change for $1\lambda_Y$

We have also studied the annual turnover for each portfolio. The momentum portfolios change about 90-100% of their stocks each period. They indicate high trading costs for these strategies. The number of stocks varies slightly and is the lowest for the negative momentum portfolio, but no lower than 31 stocks which were the low in 2020.

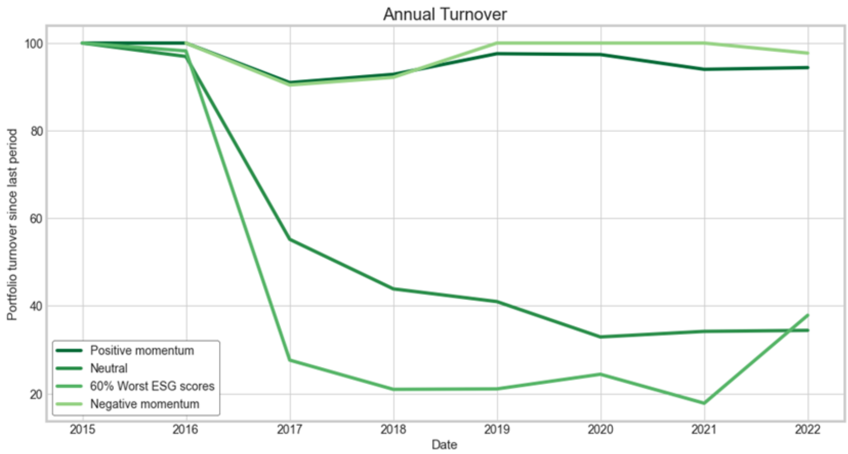

Figure 15: Annual turnover for $1\lambda_Y$

The Refinitiv data showed no excess returns over the benchmark. Regarding the ESG scores, all portfolios experienced positive adjustments. Nevertheless, the negative portfolio experienced the most positive changes, while the positive portfolio had the most negative ones. In most cases, we saw that the negative momentum portfolio had a higher SR and average return than the positive portfolio. Yet, these portfolios never achieved statistically significant positive alpha values.

6.3.2.2 ESG $3\lambda_{\mathbf{Q}}$ | BM : WORLD ESG LEADERS

Very similar results as the ESG $1\lambda_Y$ in terms of returns and scores. The World ESG leaders benchmark yielded superior returns over the period (2015-2023), outperforming the momentum portfolios in 6 out of 8 years. The negative momentum portfolio outperformed the positive. Notably, the positive portfolio, assessed through a 3-factor model, was the only one to register significant alpha at a value of -0.005411 in monthly terms. Portfolio sizes varied, with the smallest being 18 stocks for the negative momentum portfolio in 2019, though it usually consisted of around 50 stocks. The positive momentum portfolio fluctuated between 60 and nearly 500 stocks. There was a distinct trend where the positive portfolio endured the greatest ESG score reductions (-18), while the negative portfolio benefited from the most score increases (+11).

6.3.2.3 ESG $3\lambda_{\mathbf{Q}}$ | BM : EUROPE ESG LEADERS

Utilizing only EU stocks and the Europe ESG Leaders as a benchmark, both portfolios yielded cumulative returns that surpassed the benchmark. The negative momentum portfolio beat the benchmark in 6 out of 8 periods, while the positive portfolio did so in 4 out of 8 periods. Despite this, neither portfolio recorded significant alpha. The ESG score pattern was consistent with what we saw in the worldwide trend.

6.3.2.4 ESG $1\lambda_{\mathbf{Y}}$ | BM : EUROPE ESG LEADERS

This was the only instance where the positive portfolio exceeded the returns of the negative portfolio. Neither had returns good enough to generate significantly alpha returns. This marks a decrease in the consistent winning streak of the negative portfolio.

6.3.2.5 ESG $3\lambda_{\mathbf{Q}}$ | BM : USA ESG LEADERS

No portfolios recorded significant alpha values. The positive portfolios experienced the most significant downgrades, reflecting an overall decrease in ESG score of -8, while the negative portfolio saw an increase of almost +5 over the same period. These outcomes are largely consistent with the results observed previously.

6.3.2.6 ESG $1\lambda_{\mathbf{Y}}$ | BM : USA ESG LEADERS

The returns analysis revealed a substantial outperformance by the negative portfolio. However, closer inspection suggests that this superior performance, largely due to the 2020-2021 period when the negative portfolio only held 8 stocks (averaging 35 overall), may be attributed to chance. This is further supported by the absence of significant alphas for both the 3- and 5-factor model.

6.3.2.7 Other results In addition to the tests reported in this paper, we tried different combinations for the top percentile ESG scores parameter, frequencies, the minimum requirement of ESG change and different $n\lambda_i$. We did not find any new results of interest than we already had.

6.4 Momentum Trend Model

Our first step involved executing the parameter optimization on the in-sample training set. The most recent parameter optimization outcomes are presented in Appendix Table 14. Impressively, the model achieved a maximum accuracy of 63%. While this is a commendable result, we anticipate a decrease in correct predictions when the model is applied to out-of-sample data. However, it is promising that the model managed to outperform all three benchmarks, including predictions for only increases (51%), only decreases (38%), and random predictions (47%). We proceeded by utilizing the optimal parameters discovered during the training set and ran them on the out-sample dataset.

The model demonstrated an accuracy of 60% and detected a trend, on average, 47% of the time. Remarkably, it surpassed all three benchmarks, with the random prediction model achieving a correct rate of only 45%. The distribution and ECDF plot reveal that 23.9% of companies had a 100% prediction accuracy rate throughout the period.

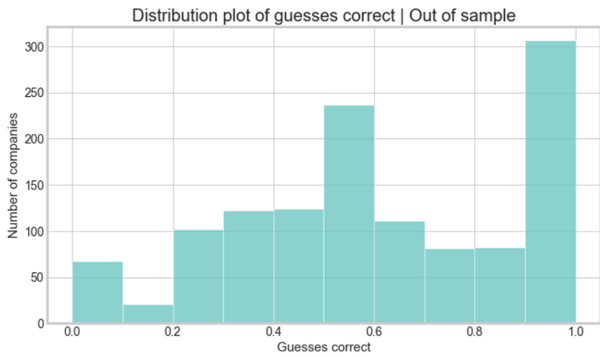

Figure 16: Guesses correct distribution

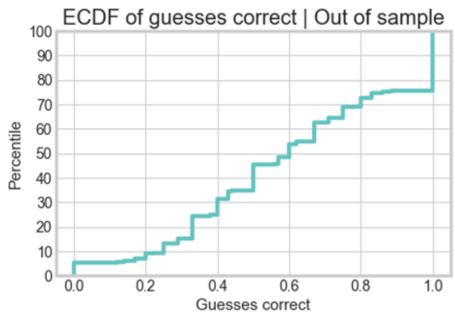

Figure 17: Guesses correct ECDF

This model tries to find trends/momentum relying on historical ESG scores. It has now been demonstrated that the model can produce a result that beats all three benchmark models - which can be seen as evidence of momentum. The absence of momentum would imply a lack of discernible trends, rendering the model no more effective than random predictions. While 60% may not be a very high accuracy, it is sufficient to claim that ESG scores show a weak form of momentum.

In an effort to eliminate some of the ”impossible” companies, we generated four additional data set. Each data set was constructed by selecting the top 50% performing companies in each category (P1: most correct predictions, P2: least square errors, P3: most correct, P3: least incorrect) from the in-sample data set. These companies were then assessed using the out-of-sample data (Appendix D). However, none of these alternative data set yielded significantly improved results, all deviating by just ±1% from the full out-of-sample test.

6.5 Analysis and discussion of the results

At the beginning of this paper, we put forth a theoretical framework involving three key drivers of momentum: company, industry and country momentum.

In order to investigate these theories, we performed a pooled regression with different λ's and fixed-effect regressions with industry and country dummies. Firstly, coefficients from the pooled regression show that there is mostly an inverse relationship between the score changes and the λ's. Indicating that company momentum does not exist. Further investigation using the Momentum Trend Model, we successfully predicted trends 100% correct for 23.9% of all the companies. Indicating that a good number of companies are showing signs of individual momentum. Secondly, the fixed-effect regression shows significant industry and country effects in almost every case. However, not all of these regressions display significant industry or country intercepts. Refinitiv's data displays no significant industry or country intercepts but weak signs of company momentum. Overall, we are confident enough to conclude that there are weak momentum tendencies in the company, industry, and country.

Is there autocorrelation present, for example, when a company improves their ESG score – is this followed by more increases in the future?

Our regression analysis reveals that ESG $1\lambda_Y$ possessed the most substantial significant coefficient at -0.22. Meaning that if a score decreases/increases by 10% in a year, on average it will change by 2.2% the subsequent year in the opposite direction. However, this observation should not be misconstrued as an indication of a lack of momentum. Rather, it suggests that after an initial shift in any direction, there is a partial reversion in the opposite direction during the next period. Moreover, we began to observe significant positive coefficients forming after $4\lambda_Y$, suggesting potential momentum over higher timeframes.

Are there systemic effects in changes in ESG scores?

Utilizing the Momentum Quantile model, we found that portfolios containing companies with the highest ESG score declines (Q5) exhibited the most positive adjustments during the holding period. While portfolios with companies exhibiting the most ESG growth (Q1) in the previous period experienced the greatest score decreases. Which was expected and in line with the regression results. Interestingly, the model displayed a disturbance in the systematic pattern when applied with an insignificant λ_Q. Implying that the systematic pattern may only be preserved at the significant frequency identified through regression analysis. However, this finding is subject to chance as we did not have sufficient time to test this across all yearly frequencies. Analyzing the ESG cumulative changes in Figure 8, we noticed that Q4 and Q5 exhibited approximately double the cumulative score change compared to Q1 and Q2 in the opposite direction. This observation is particularly intriguing as it suggests companies respond more aggressively to significant decreases in score, working harder to enhance their scores after a substantial decline. Meanwhile, firms that experience an increase in their scores face a comparatively minor decrease.

A plausible explanation for the negative coefficient and the reversion pattern could be linked to the scoring methodology of the MSCI. The score can shift if any of the underlying Environmental, Social, or Governance (ESG) pillars breach certain thresholds, thereby pushing for a change in the overall score. A significant portion of the data used to evaluate companies is sourced from quarterly and annual reports, which often reflect a certain level of seasonality or at least vary from one report to another. If these data points hover close to the thresholds necessary for a score change, it can lead to score volatility, making them noisy. If a company suddenly gets a lot of data points barely within the scoring bounds, it might be more likely that more fall outside again next period. This implies that the more the score increases i.e. more data points fall within the boundaries - the higher the likelihood of some reverting back outside. This dynamic could possibly explain the systematic patterns we have observed where the biggest changes also lead to the highest reversions.

Assuming this reasoning explains score decrease for firms that previously exhibited significant increases, we can further assume that this holds equally true for companies that register an upward score adjustment after a notable decline in the previous period. This could be perceived as a natural "fall back" that occurs following a substantial score adjustment. However, as we discovered earlier, the negative reaction is approximately twice as strong as the positive reaction, which leaves about 50% of the score increase unexplained.

Many investors consider ESG an essential factor in investment decisions and are usually important to the companies themselves. A steep decline in ESG scores can result in negative reactions from stakeholders or the board, placing the company under pressure to improve. This could prompt intensified ESG focus and efforts, leading to an increase in their score in the subsequent period. It is also not unusual that incidents and controversies like data breaches, labor controversies, or environmental incidents can raise and get flagged with the scoring company resulting in a decrease in score. However, if the company takes swift and meaningful action to address the issue, this could lead to a rebound in the subsequent period.

Can we identify trends in these adjustments?

So far, our analysis has uncovered short-term negative autocorrelation that transitions into a positive relationship over the long term. We've also confirmed this can be utilized to detect systematic effects in ESG scores. However, the crucial question persists: does momentum exist within ESG scores?

When considering a single company, the evidence is insufficient to confirm the existence of momentum within the scores. Nonetheless, by constructing portfolios where we buy/sell companies that decreased/increased the most in scores during the last year, hold them for a year, and then rebalance, you will in fact have created a positive ESG momentum portfolio. But this might be considered cheating, therefore we needed a third model that utilized a more complex structure to identify trends. If we can find trends and then use them to predict future trends - we can say with more confidence that momentum does exist in ESG scores.

The "Momentum Trend Model" does precisely that and provides the first evidence of real momentum as we typically know it from finance –a longer and sustainable trend. We see the model predicts pretty accurately for many companies, but also, many are entirely unpredictable with this method. A predicting accuracy of 60% is not good enough to call yourself a clairvoyant but we think it is enough to conclude ESG scores show signs of weak form of momentum.

A simple yet efficient model for predicting ESG score fluctuations would involve making guesses contrary to the previous period's change, aligning with the negative 1-year change coefficient. We, however, opted not to do this and instead developed a trend model. Simply predicting whether a score will rise or fall in the next period would lack economic insight as it does not necessarily reflect whether a score is in an upward or downward trend, but instead oscillates within the trend. Therefore, The trend model allows us to monitor long-term trends that are more likely to provide valuable data. This aspect can be expanded upon, for instance, to explore whether holding companies exhibiting uptrends yield higher returns rather than attempting to exploit minor shifts without considering the overall trend.

The potential value of predicting scores and creating momentum portfolios

Our "Momentum Strategy Model" findings indicated that the negative momentum portfolio consistently witnessed the most substantial score increases and generally yielded the highest returns. Conversely, the positive momentum portfolio endured the most significant score reductions during the holding period and frequently underperformed. This pattern suggests that ESG score increases are associated with higher returns and decreases with lower returns - albeit with almost no statistically significant alpha against the benchmarks. This is only the case for ESG leaders, whereas we find no relationship using the Quantile model that does not isolate leaders. These findings align with the observations made by Shanaev & Ghimire (2022) mentioned earlier in terms of adjustments and returns and that the effect is bigger when considering only ESG leaders. It's worth noting that the insignificant alpha values we observed could potentially be due to the utilization of value-weighted benchmarks.

We have seen that by creating portfolios that accumulate on these small up ticks of scores, which could be a natural and somewhat random effect rather than a direct result of company actions, has shown no significant impact on returns. The relationship observed was mildly positive at most. This suggests that occasional fluctuations are of little importance for investors. For instance, an investor is unlikely to react significantly to a change in a company's score from 85 to 84. However, a consistent downward trend, say from 85 to 84, 80, 76, and then to 69, is more likely to provoke concern and prompt investor action. A bigger-picture trend is more likely to trigger actions from investors. We believe this could explain the lack of alpha return like others before us have found (Bekaert et al., 2023), (Dimson et al., 2015).

Evidence against the theory

Our research contrasts with Sankar et al. (2019), who report that the positive momentum portfolio outperformed using European companies. However, this was the only instance where the positive portfolio surpassed the negative in our tests. This might have been an outline of European companies reacting differently to score changes. A crucial critique of their study is the absence of investigation into the changes in scores during the holding period. Our study indicated that their positive momentum portfolio is likely experiencing on average score reductions. This suggests that negative score adjustments yield higher returns. Furthermore, Sankar et al. (2019) did not evaluate significance of alpha but rather simply compared cumulative returns. Considering all these factors, we maintain that the evidence provided is not sufficiently convincing to dismiss the theory of a positive relationship between score increases and returns.

Additional findings

For the yearly MSCI throughout the regression analysis, there is a pattern that the governance pillar score's λ's tends to have the largest coefficient of the pillars. Significantly higher than the other pillars λ's and in the same order of magnitude as the ESG scores λ's coefficient. This is consistent with what we find in chapter 2.2 from Crespi & Migliavacca (2020), the governance pillar follows a different and very strong trend which seems to drive the overall increase in the ESG score over time. This can potentially be connected to the fact that the governance pillar can be viewed as a measure of the quality of management in firms. Hence shareholders will always be strictly evaluating this particular pillar. Therefore, it may be of special importance and be connected over time. We also agree with their findings that the governance pillar react in an opposite way to the country factor, compared to the environmental and social pillar. The economic reasoning behind this result is that most countries are liberal on diversity and corporate governance, leading to lose regulations, henceforth a negative impact on the governance pillar score. As (Ehlers et al., 2022) finds that the environmental pillar lack of reporting, in their study of Refinitiv ESG scores. This could be one explanation for the weak explanatory power of the environmental pillar's λ's on the score change. Which we observe in the results from the pooled regression for both the Refinitiv data set where there are no significant λ's and the yearly frequency for the MSCI data set where only $7\lambda_Y$ is significant. However, within the context of industry effects, the environmental pillar appears to be the most significant and important pillar. This suggests that differences across industries are more susceptible to the impacts of emission regulations.

7 Conclusion

This paper explored the notion of momentum within Environmental, Social, and Governance (ESG) scores, investigating three potential drivers of momentum: company, industry, and country momentum. Contrary to the initial expectation, the regression analysis found no strong evidence to support the existence of company momentum. However, further exploration through a Momentum Trend Model indicated that a good number of companies showed signs of individual momentum over longer periods, though its predictive accuracy is still limited. Further, the industry and country effects exhibited in our regression analyses suggest the existence of momentum on these broader scales, even if the strength of this momentum is generally weak.

The findings from our regression analysis suggested an interesting aspect of score dynamics and autocorrelation. Showing a significant negative coefficient of -0.22 when regressing one-year change in ESG score on $1\lambda_Y$ (lagged one-year change). However, from $5\lambda_Y$ onwards, we observed significant but low positive coefficients, indicating potential momentum over longer periods. This reversion pattern may be tied to MSCI's scoring methodology, which is influenced by thresholds within the ESG pillars. The data, often from quarterly and annual reports, can hover near these thresholds, causing score volatility. If many data points fall just within scoring bounds, some may fall outside in the next period. This means that higher score changes could lead to higher chance of reversion. This dynamic could explain the observed systematic patterns and the "fall back" after a large score adjustment.

Our investigation into systemic effects through a ranking model shows that companies with the largest ESG score declines (Q5) made the most positive adjustments during the holding period. Meanwhile, those with significant ESG growth (Q1) experienced the greatest score decreases. In studying ESG cumulative changes, Q4 and Q5 saw about double the cumulative score change in the opposite direction compared to Q1 and Q2. This suggests that companies respond more aggressively to substantial score declines, striving harder to improve their scores. The negative reaction is about twice as strong as the positive reaction, leaving about 50% of the score increase unexplained. This could be due to negative reactions from investors or boards, creating pressure to improve their ESG scores after a steep decline. Additionally, incidents and controversies can result in score decreases, but swift and effective action can cause a rebound in the subsequent period.

The potential value of predicting scores and creating momentum portfolios was examined through our "Momentum Strategy Model". The model indicated that portfolios with negative momentum tended to have more significant score increases and yielded higher returns. Conversely, portfolios with positive momentum frequently experienced score reductions and underperformed. But no portfolios recorded any significant risk adjusted alpha. This result was consistent with previous findings by Shanaev & Ghimire (2022) and differed from the conclusions drawn by Sankar et al. (2019), who reported outperformance by the positive momentum portfolio.

In conclusion, we reject the null hypothesis, "that there exists no momentum in ESG scores, the ESG scores are random and do not exhibit a trend over time". We can say that there are systemic effects in changes in ESG scores. Autocorrelation is present, for example, when a company experience a decrease in their ESG score – it is usually followed by a smaller increase in following year. We identify about 60% of these trends but do not find any significant alpha. Our research has demonstrated that while evidence of momentum exists within ESG scores, it is not consistently strong across all companies.

Our study has contributed to the growing body of literature examining the dynamics of ESG scores and their implications. While the results provide intriguing insights, they also highlight areas for further research. Future studies may wish to explore the variations in ESG score momentum across different industries and countries further and investigate the impact of different ESG pillars have on score momentum. They might also consider delving deeper into the mechanisms driving the reversion pattern observed in score changes. Perhaps the most interesting would be to create an investment strategy based on the longer score trends identified through the Momentum Trend model. However, to effectively operationalize such a strategy, one would likely need to enhance the accuracy of the predictions.

Hopefully, this research will spark further investigation into this area, contributing to a more nuanced understanding of ESG score dynamics and their implications for investors and companies alike. As ESG scores continue to play a pivotal role in investment decisions, this line of research will undoubtedly remain of significant interest to academics, investors, and corporations.

Our study is also limited as we have worked with quite a short time series. Ultimately, exploiting the longer trends with the momentum trend model is challenging since we have to split the data set into two subsamples. One can anticipate improvement in the quality of ESG data, together with an extended time series. Consequently, this will enable a more robust exploration of our findings, offering opportunities to investigate potential longer-term momentum.

APPENDIX

A Figures from Data and Analysis

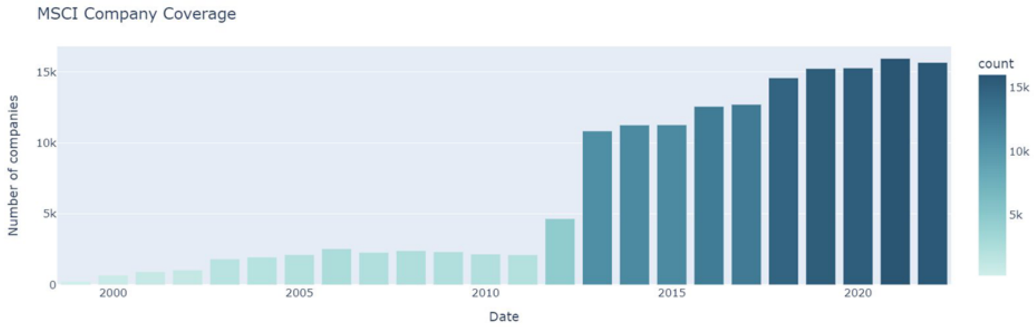

Figure 18: MSCI Company Coverage

The figure shows the number of companies covered by the MSCI data set in each year.

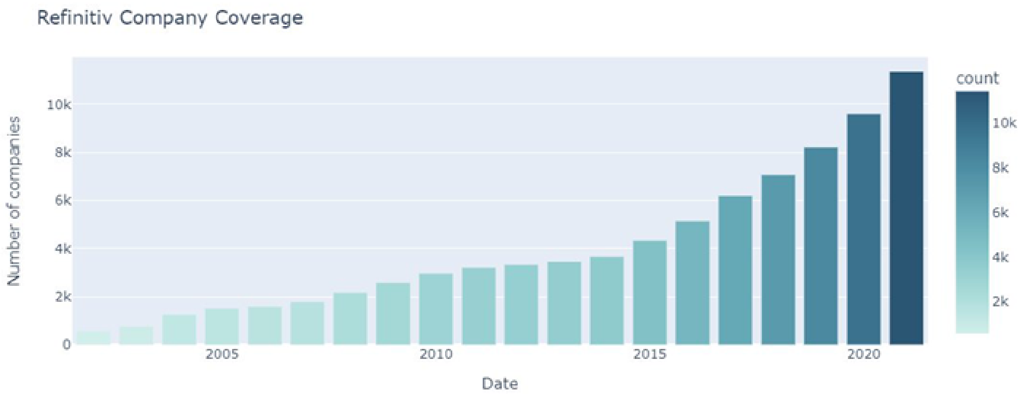

Figure 19: Refinitiv Company Coverage

The figure shows the number of companies covered by the Refinitiv data set in each year.

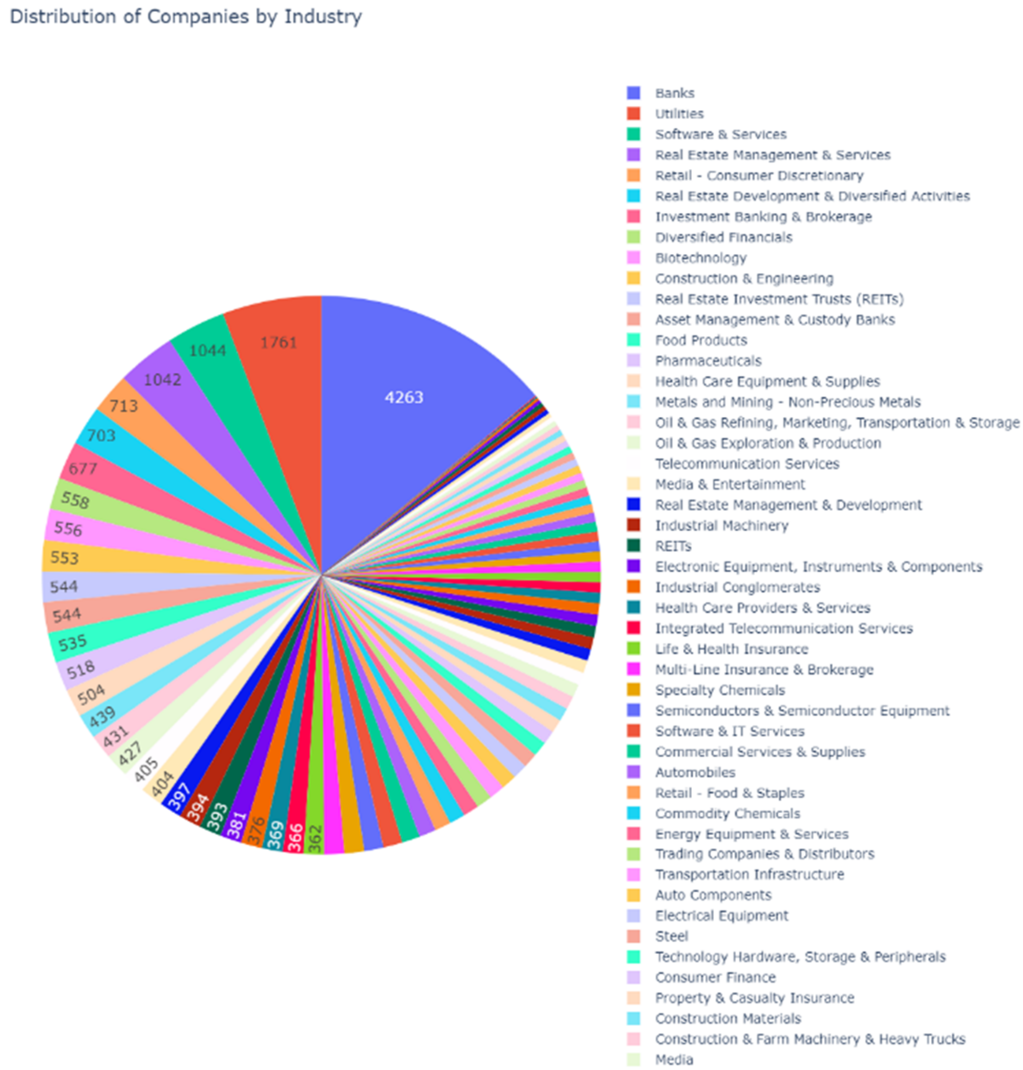

Figure 20: Distribution of Industries - MSCI

The figure shows the distribution of industries within the MSCI data set.

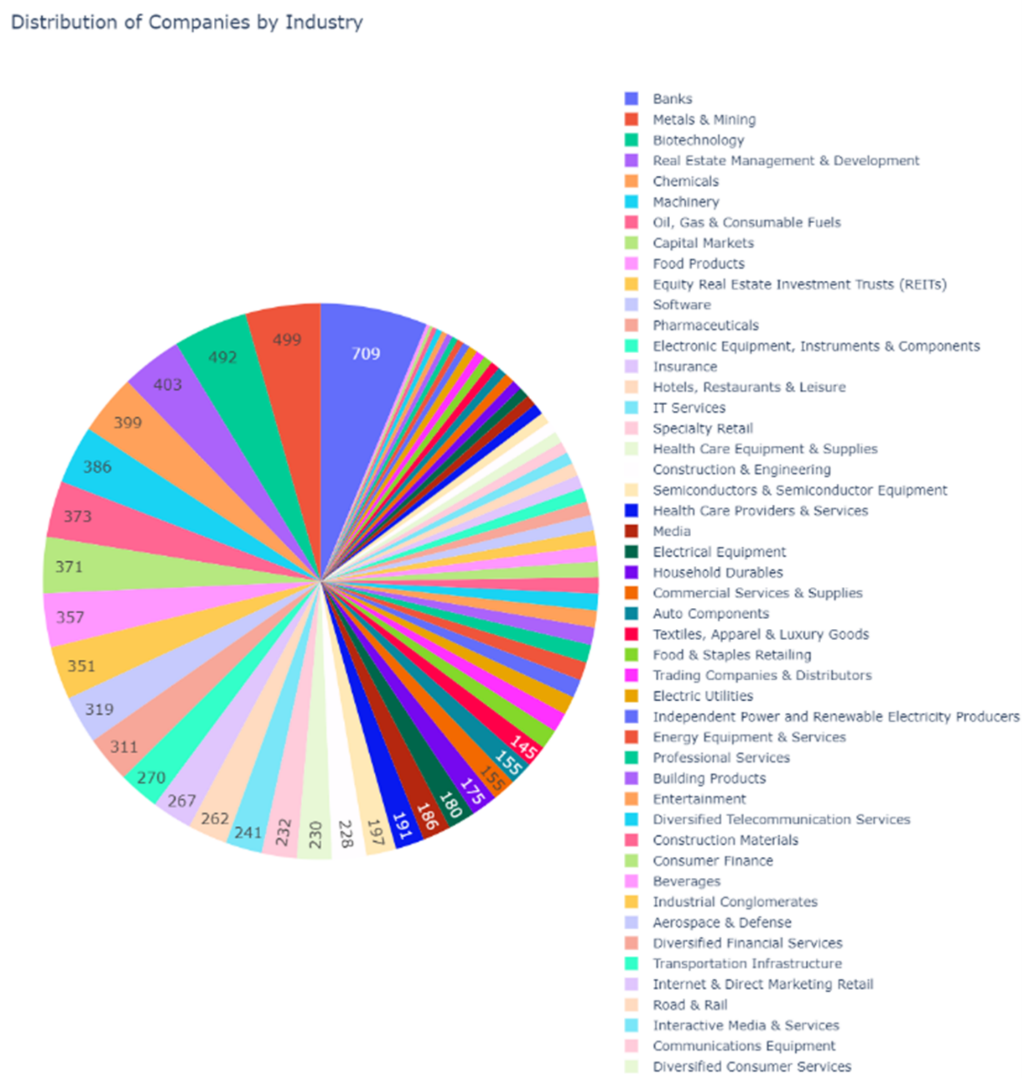

Figure 21: Distribution of Industries - Refinitiv

The figure shows the distribution of industries within the Refinitiv data set.

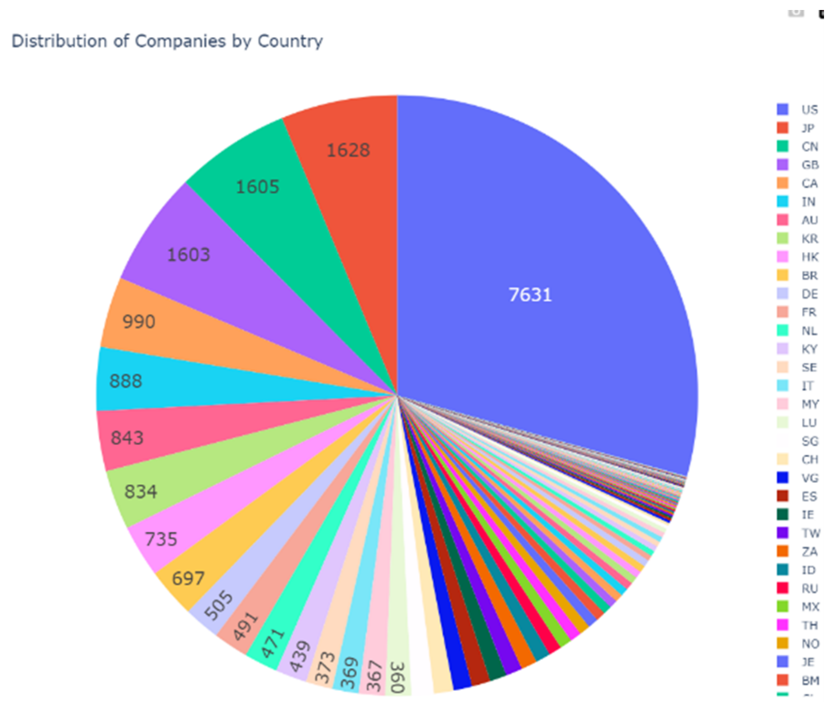

Figure 22: Distribution of Countries - MSCI

The figure shows the distribution of countries within the MSCI data set.

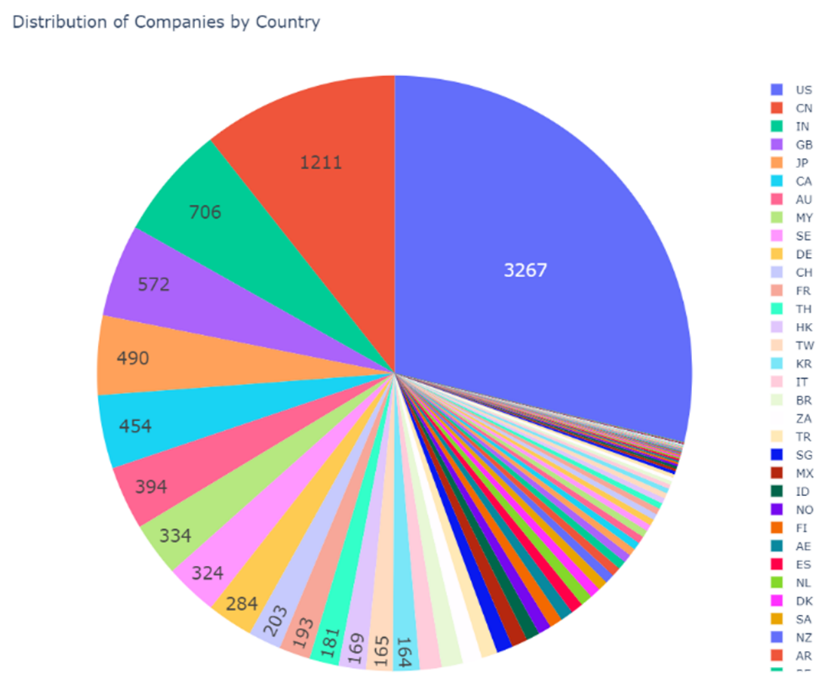

Figure 23: Distribution of Countries - Refinitiv

The figure shows the distribution of countries within the Refinitiv data set.

B Tables from Regression Analysis

Table 7: F-test - Results

Data set	Frequency	Pillar	Pooled Regression	Fixed-Effect Regression	
			F-stat	Industry F-stat	Country F-stat
MSCI	Quarterly	ESG	80.86*	23.71*	
		E	19.49*	12.14*	
		S	21.63*	24.90*	
		G	64.95*	4.67*	
MSCI	Yearly	ESG	40.00*	157.21*	22.57*
		E	2.97*	35.05*	27.91*
		S	5.08*	75.26*	19.26*
		G	8.26*	36.83*	36.77*
Refinitiv	Yearly	ESG	1.89*	17.77*	18.76*
		E	0.047	5.00*	5.25*
		S	3.61*	15.27*	10.26*
		G	4.60*	19.99*	17.32*

This table shows the F-statistic from the F-test of each regression, which includes the pooled regressions and fixed-effect regressions at each frequency and pillar. Note: * means that the test is significant at the 5% significance level.

Table 8: Pooled Regression

Data-set	Frequency	Pillar	Significant λ	Coefficient	p-value
MSCI	Quarterly	ESG	Intercept	0.0079	0.000
			3	-0.0186	0.000
			4	-0.0150	0.000
		E	Intercept	0.0195	0.000
			2	0.0043	0.000
			3	0.0041	0.002
			4	-0.0087	0.000
		S	Intercept	0.0181	0.000
			4	-0.0105	0.000
		G	Intercept	0.0270	0.000
			2	-0.0067	0.000
			3	-0.0081	0.001
			4	-0.0125	0.000
MSCI	Yearly	ESG	Intercept	0.0199	0.000
			1	-0.2199	0.000
			2	-0.0527	0.001
			5	0.0465	0.000
			6	-0.0332	0.000
			7	0.0120	0.000
		E	Intercept	0.0906	0.000
			7	-0.0051	0.002
		S	Intercept	0.0582	0.000
			1	-0.0364	0.001
			5	-0.0045	0.003
			7	-0.0063	0.000
		G	Intercept	0.0235	0.000
			1	-0.1999	0.000
			4	-0.0693	0.000
Refinitiv	Yearly	ESG	Intercept	0.0771	0.000
			1	-0.0814	0.000
			2	0.0568	0.000
		E	Intercept	0.5347	0.002
		S	Intercept	0.0890	0.000
			1	-0.0321	0.001
			2	0.0478	0.000
		G	Intercept	0.1290	0.000
			3	-0.0651	0.000
			4	0.0292	0.000

The table shows the results from the pooled regression for the overall ESG score and the three pillar (E,S,G) score for all frequencies, which includes the regression coefficient and p-value for each significant lambda in each regression.

Table 9: Industry-fixed Regression - Quarterly Frequency - MSCI

		Industry	
Pillar	Significant λ	Coefficient	p-value
ESG	Intercept	0.0065	0.000
	3	-0.0188	0.000
	4	-0.0174	0.000
E	Intercept	0.0741	0.000
	2	0.0041	0.001
	3	0.0040	0.002
	4	-0.0099	0.000
S	Intercept	0.0139	0.000
	4	-0.0130	0.000
G	Intercept	0.0275	0.000
	2	-0.0068	0.000
	3	-0.0083	0.000
	4	-0.0128	0.000

The table shows the results from the industry-fixed regression for the overall ESG score and the three pillar (E,S,G) scores for quarterly frequency in the MSCI data set. The results includes the regression coefficients and p-values of each significant lambda in each regression.

Table 10: Significant industries - Quarterly Frequency - MSCI

ESG			E			S			G		
Industry	Coef	p-value	Industry	Coef	p-value	Industry	Coef	p-value	Industry	Coef	p-value
Energy Equip. Services	0.0066	0.000	Wireless Telecom. Services	-0.0677	0.002	Media Entertain.	0.0095	0.004	Trading Companies Distrib.	-0.0168	0.039
Oil Gas Explor. Prod.	0.0060	0.000	Media Entertain.	-0.0629	0.003	Energy Equip. Services	0.0104	0.003	Integr. Telecom. Services	-0.0203	0.007
Food Prod.	0.0037	0.004	Energy Equip. Services	-0.0584	0.005	Oil Gas Explor. Prod.	0.0405	0.000	Life Health Insurance	-0.0208	0.000
Auto Comp.	-0.0071	0.000	Retail Consumer Discret.	-0.0631	0.004	Auto Comp.	-0.0119	0.005	Investm. Banking Brokerage	0.0155	0.005
Specialty Chemicals	0.0093	0.000	Oil Gas Explor. Prod.	-0.0686	0.002	Specialty Chemicals	0.0473	0.000	Casinos Gaming	0.0376	0.020
Utilities	0.0060	0.000	Electr. Equip. Instrum. Comp.	-0.0736	0.001	Utilities	0.0266	0.000	Diversif. Chemicals	-0.0180	0.003
Restaur.	-0.0055	0.007	Auto Components	-0.0743	0.001	Restaur.	-0.0102	0.001	Broadcast. Cable Satellite	-0.0172	0.003
REITs	-0.0105	0.000	Paper Forest Products	-0.0627	0.004	REITs	-0.0446	0.000	Health Care Tech.	0.0915	0.000
RE Mngmnt Services	0.0069	0.000	Construct. Materials	-0.0633	0.003	RE Mngmnt Services	0.0472	0.000			
Telecom. Services	0.0080	0.000	Restaur.	-0.0813	0.001	Integr. Telecom. Services	0.0132	0.004			
Auto-mobiles	-0.0113	0.000	RE Mngmnt Services	-0.0621	0.004	Banks	-0.0152	0.000			
Health Care Providers Services	0.0055	0.000	Trading Companies Distrib.	-0.0685	0.002	Constr. Engineer.	0.0305	0.000			

Construct. Engineer.	0.0064	0.000	Integr. Telecom. Services	-0.0687	0.002	Pharma.	-0.0106	0.011
Electrical Equipment	-0.0038	0.005	Telecom. Services	-0.0684	0.002	Casinos Gaming	-0.0096	0.020
Casinos Gaming	0.0075	0.000	Tech. Hardware Storage Peripherals	-0.0709	0.002	Constr. Farm Machinery Heavy Trucks	0.018	0.010
RE Develop. Diversif. Act.	0.0124	0.000	Health Care Equip. Supplies	-0.0627	0.003	RE Develop. Diversif. Act.	0.0783	0.000
Oil Gas Refining Marketing	-0.0096	0.000	Software Services	-0.0587	0.005	Oil Gas Refining Marketing	0.0449	0.000
Metals & Mining Non-precious Metals	0.0063	0.000	Life-Health Insurance	-0.0617	0.004	Proff. Services	0.0108	0.050
Semicond. Equip.	0.0052	0.001	Auto-mobiles	-0.0737	0.001	Transport. Infrastr.	0.0213	0.000
Containers Packaging	-0.0057	0.001	Health Care Providers Services	-0.0694	0.002	Industr. Conglom.	0.0170	0.020
Beverages	0.0097	0.000	Construct. Engineer.	-0.0731	0.001	Steel	-0.0117	0.001
RE Mngmnt Develop.	-0.0062	0.001	Biotech.	-0.0596	0.004	Air Freight Logistics	0.0371	0.004
Prof. Services	0.0056	0.001	Electric. Equip.	-0.0727	0.001	Integr. Oil Gas	0.0207	0.000
Transport. Infrastr.	0.0070	0.001	Industr. Machinery	-0.0783	0.000	Health Care Tech.	-0.0209	0.004
Metals & Mining Precious Metals	0.0087	0.000	Construct. Farm Machinery Heavy Trucks	-0.0715	0.002			

Interact. Media Services	0.0129	0.001	RE Develop. Diversif. Act.	-0.0628	0.003			
Tobacco	0.0429	0.000	Oil Gas Refining Marketing	-0.0632	0.003			
Com. Equipment	-0.0195	0.005	Building Products	-0.0700	0.002			
Broadcast. Cable Satellite	-0.0051	0.004	Home Building	-0.0721	0.002			
Health Care Tech.	0.0493	0.000	Semicon. Equip.	-0.0640	0.003			
			Contain. Packaging	-0.0641	0.004			
			Hotels Travel	-0.0622	0.004			
			Retail Food Staples	-0.0682	0.002			
			Diversif. Consumer Services	-0.0595	0.005			
			Aerospace Defense	-0.0786	0.001			
			Consumer Finance	-0.0630	0.003			
			Proff. Services	-0.0593	0.005			
			Household Personal Products	-0.0731	0.001			
			Interact. Media Services	-0.0624	0.004			
			Commodity Chemicals	-0.0616	0.004			
			Industrial Conglom.	-0.0811	0.001			
			Marine Transport	-0.0648	0.003			
			Diversif. Chemicals	-0.0616	0.002			

Integr. Oil Gas	-0.0630	0.003		
Com. Equip.	-0.0769	0.004		
Broadcast. Cable Satellite	-0.0658	0.003		
Health Care Tech	-0.0686			

The table shows significant industries from the industry-fixed regression for the quarterly frequency for the overall ESG score and the three pillar (E,S,G) scores in the MSCI data set, the results include the industries coefficients and p-values.

Table 11: Fixed Effect Regression Yearly - MSCI

				Industry		Country	
Data set	Frequency	Pillar	Significant λ	Coefficient	p-value	Coefficient	p-value
MSCI	Yearly	ESG	Intercept			-0.0065	0.000
			1	-0.2648	0.000	-0.2218	0.000
			2	-0.0718	0.000	-0.0532	0.000
			5	0.0475	0.000	0.0472	0.000
			6	-0.0325	0.000	-0.0336	0.000
			7	0.0082	0.001	0.0119	0.000
		E	Intercept			-0.0751	0.000
			7	-0.0071	0.000	-0.0051	0.017
		S	Intercept			-0.0054	0.005
			1	-0.0554	0.000	-0.0395	0.001
			5	-0.0056	0.001	-0.0048	0.003
			7	-0.0071	0.000	-0.0065	0.000
		G	1	-0.2184	0.000	-0.2193	0.000
			4	-0.0702	0.000	-0.0713	0.000
			6			0.0043	0.003
Refinitiv	Yearly	ESG	1			-0.0735	0.007
			2			0.0492	0.014
		E					
		S	1	-0.0388	0.020	-0.0322	0.013
			2			0.0474	0.003
		G	Intercept	0.1120	0.019		
			3	-0.0636	0.005	-0.0621	0.003
			4	0.0286	0.000	0.0293	0.001
			5	0.-0.0197	0.017	-0.0201	0.048

The table shows the results from the fixed effect regression for the overall ESG score and the three pillar (E,S,G) scores for yearly frequency in both MSCI and Refinitiv data set, including the regression coefficient and p-values of each lambda in each regression, excluding the significant industries and countries in each regression.

Table 12: Significant industries - Yearly Frequency

	ESG			E			S	G
Data Set	Industry	Coef	p-value	Industry	Coef	p-value		
MSCI	Restaur.	-0.1314	0.001	Restaur.	-0.3059	0.000		
	Household Personal Prod.	-0.0828	0.005	Banks	0.1915	0.000		
	Building Prod.	-0.0837	0.004	Household Personal Prod.	-0.1391	0.001		
	Airlines	-0.0913	0.003					
	Media	-0.1308	0.003					
Refinitiv								

The table shows significant industries, their regression coefficient and p-value for overall ESG score and the three pillar (E,S,G) scores for the yearly frequency in both data sets. Notice that the Refinitiv data set, does not have any significant industries for the yearly frequency.

Table 13: Significant countries - Yearly Frequency - MSCI

ESG			E			S			G		
Country	Coef	p-value	Country	Coef	p-value	Country	Coef	p-value	Country	Coef	p-value
US	0.0272	0.000	US	0.1590	0.000	US	0.0713	0.000	CH	-0.0580	0.000
CA	0.0418	0.000	CA	0.1785	0.000	CA	0.0647	0.000	JP	0.0975	0.000
FI	0.0362	0.001	FI	0.1453	0.000	HK	0.0812	0.000	AU	0.0239	0.002
HK	0.0292	0.000	HK	0.1660	0.000	FR	0.0450	0.000	SE	-0.0271	0.002
BE	0.0287	0.001	BE	0.1288	0.000	CH	0.0676	0.000	TH	0.1161	0.003
FR	0.0218	0.000	FR	0.1241	0.000	JP	0.0285	0.000	IN	-0.0869	0.001
CH	0.0241	0.000	CH	0.1044	0.000	GB	0.0541	0.000	TR	-0.1439	0.002
JP	0.0257	0.000	JP	0.1163	0.000	SE	0.0536	0.000	NO	-0.0334	0.002
AU	0.0441	0.000	MA	0.3202	0.024	DE	0.0494	0.000	PL	-0.1324	0.000
GB	0.0319	0.000	AU	0.1406	0.000	CN	0.1918	0.003	ES	-0.0823	0.000
SE	0.0225	0.000	GB	0.1985	0.000	BM	0.1501	0.000	AT	-0.0561	0.002
SG	0.0180	0.005	SE	0.1308	0.000	MY	0.0470	0.000	QA	-0.2396	0.003
DE	0.0280	0.000	SG	0.1546	0.000	IN	0.1139	0.000	AR	-0.1195	0.000
CN	0.0421	0.000	TW	0.1540	0.000	BR	0.1024	0.004	GI	-0.05556	0.000
BM	0.0695	0.000	CN	0.1600	0.000	KR	0.0786	0.000	HR	-0.0935	0.004
MY	0.0367	0.000	TH	0.2103	0.008	ZA	0.1040	0.001	JE	0.2947	0.001
NO	0.0456	0.000	CZ	0.1582	0.002	NL	0.0973	0.000	PG	0.3216	0.000
BR	0.0351	0.000	MY	0.2861	0.000	MX	0.1035	0.000	GE	-0.1176	0.000
DK	0.0490	0.000	PR	0.3736	0.029	ES	0.0834	0.000	BS	-0.1025	0.003
ZA	0.0480	0.000	IN	0.1276	0.000	RU	0.1763	0.004	NA	-0.1040	0.000
NL	0.0310	0.000	IT	0.1059	0.000	SK	-0.0856	0.002	MO	0.1505	0.000
MX	0.0399	0.001	NO	0.1369	0.000	GI	0.0109	0.000	CR	-0.0872	0.000
ES	0.0182	0.003	IL	0.3967	0.001	PG	0.3120	0.000	BW	-0.1258	0.000
RU	0.0536	0.001	GR	0.1468	0.027	IM	0.2678	0.001	VE	0.2563	0.000
NZ	0.0370	0.000	BR	0.1321	0.000	GE	-0.1091	0.000	BF	-0.3370	0.000
KZ	-0.1445	0.000	IE	0.2005	0.000	BS	0.0987	0.004	BB	0.3188	0.000
GI	-0.0154	0.000	ID	0.2181	0.024	VG	0.1449	0.000	AW	0.1860	0.000
JE	0.0650	0.002	DK	0.1416	0.000	BW	-0.1909	0.000			
PG	0.1611	0.000	KR	0.1284	0.000						
IM	0.1531	0.000	ZA	0.1220	0.000						
GE	-0.0788	0.000	NL	0.1020	0.000						
VG	0.0698	0.000	MX	0.1303	0.000						
BW	-0.1342	0.000	PL	0.2076	0.000						
BF	-0.0681	0.000	ES	0.1365	0.000						
BB	0.1278	0.001	PT	0.1581	0.008						
AW	0.0157	0.000	PH	0.1004	0.035						
			AT	0.2010	0.000						

CO	0.2249	0.000
CL	0.1580	0.002
NZ	0.1025	0.000
KY	0.1997	0.000
SK	0.4014	0.000
HU	0.1351	0.048
MT	0.1572	0.002
LU	0.1171	0.000
GI	0.0725	0.000
HR	0.5765	0.000
JE	0.1394	0.003
PG	0.1211	0.000
IM	0.1416	0.002
GE	0.3032	0.000
CW	0.2427	0.043
VG	0.2083	0.000
MO	0.1291	0.022
CR	0.2197	0.000
BW	-0.1550	0.000
VE	-0.1085	0.048
BF	0.0626	0.001
BB	0.1839	0.000

The table shows significant country intercepts for the yearly frequency for all three pillar scores as well as the overall score in the MSCI data set.

C Momentum strategy model

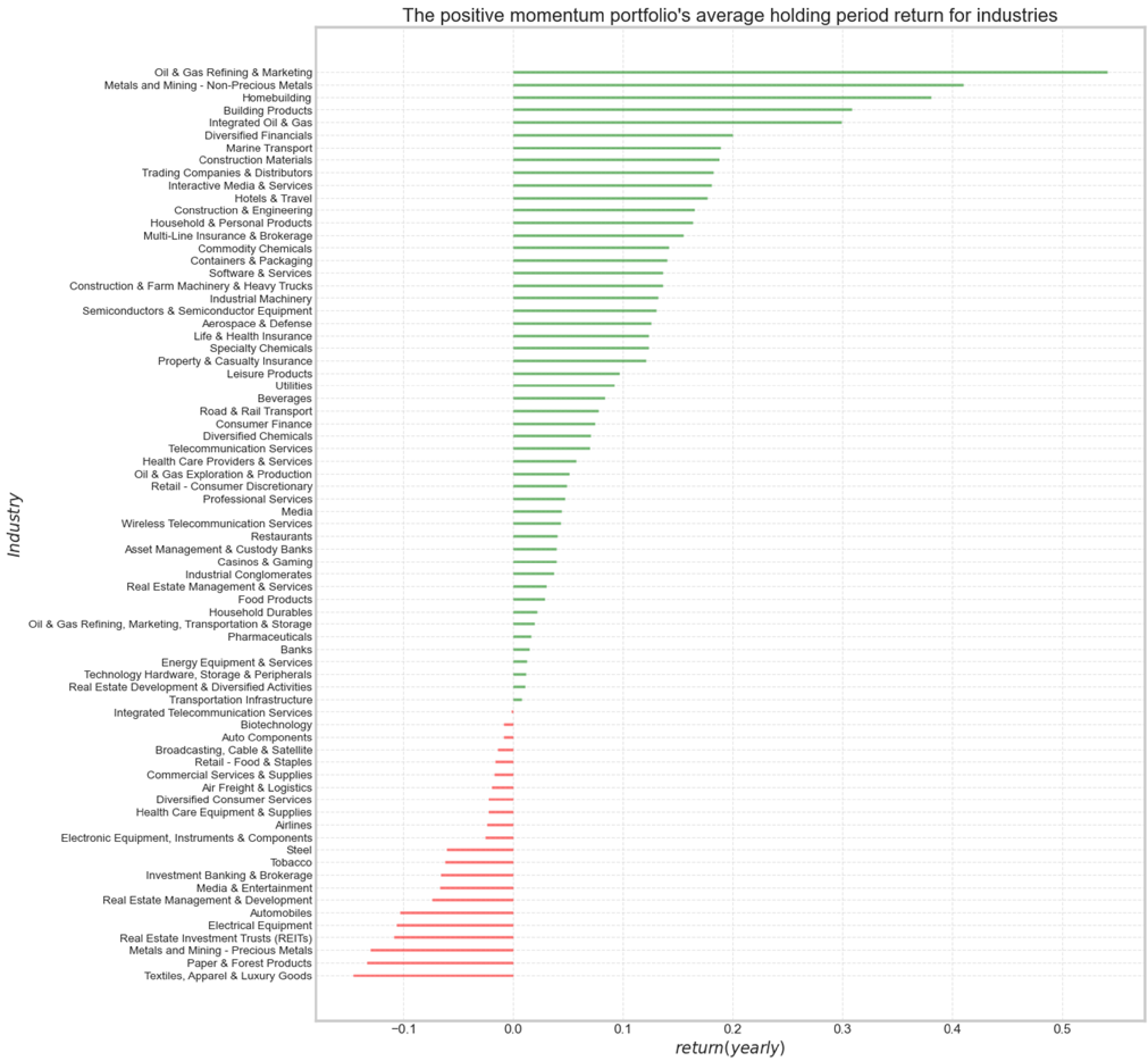

Figure 24: Positive momentum average HPR for industries, using $1\lambda_Y$

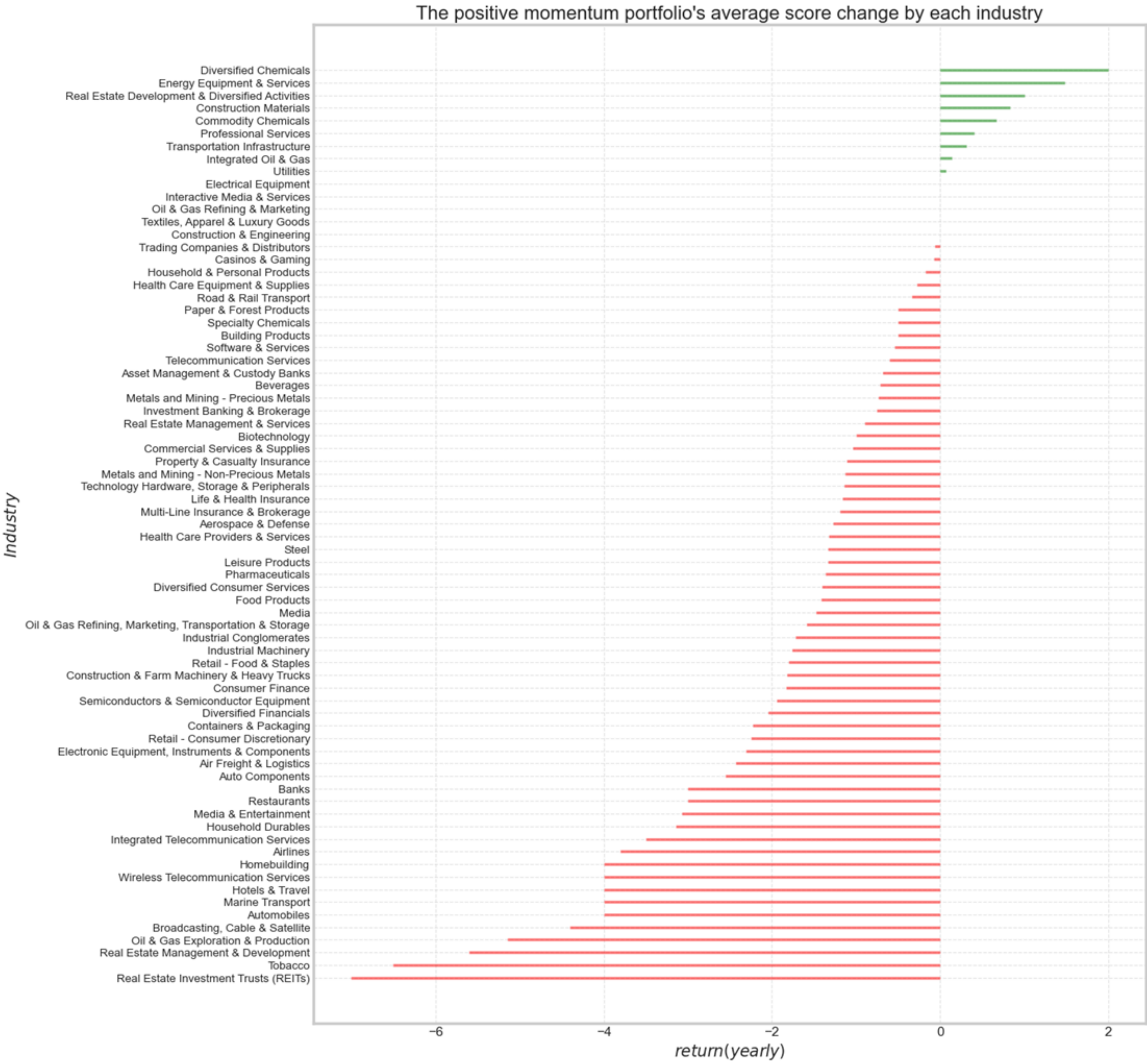

Figure 25: Positive momentum average score change for industries, using $1\lambda_Y$

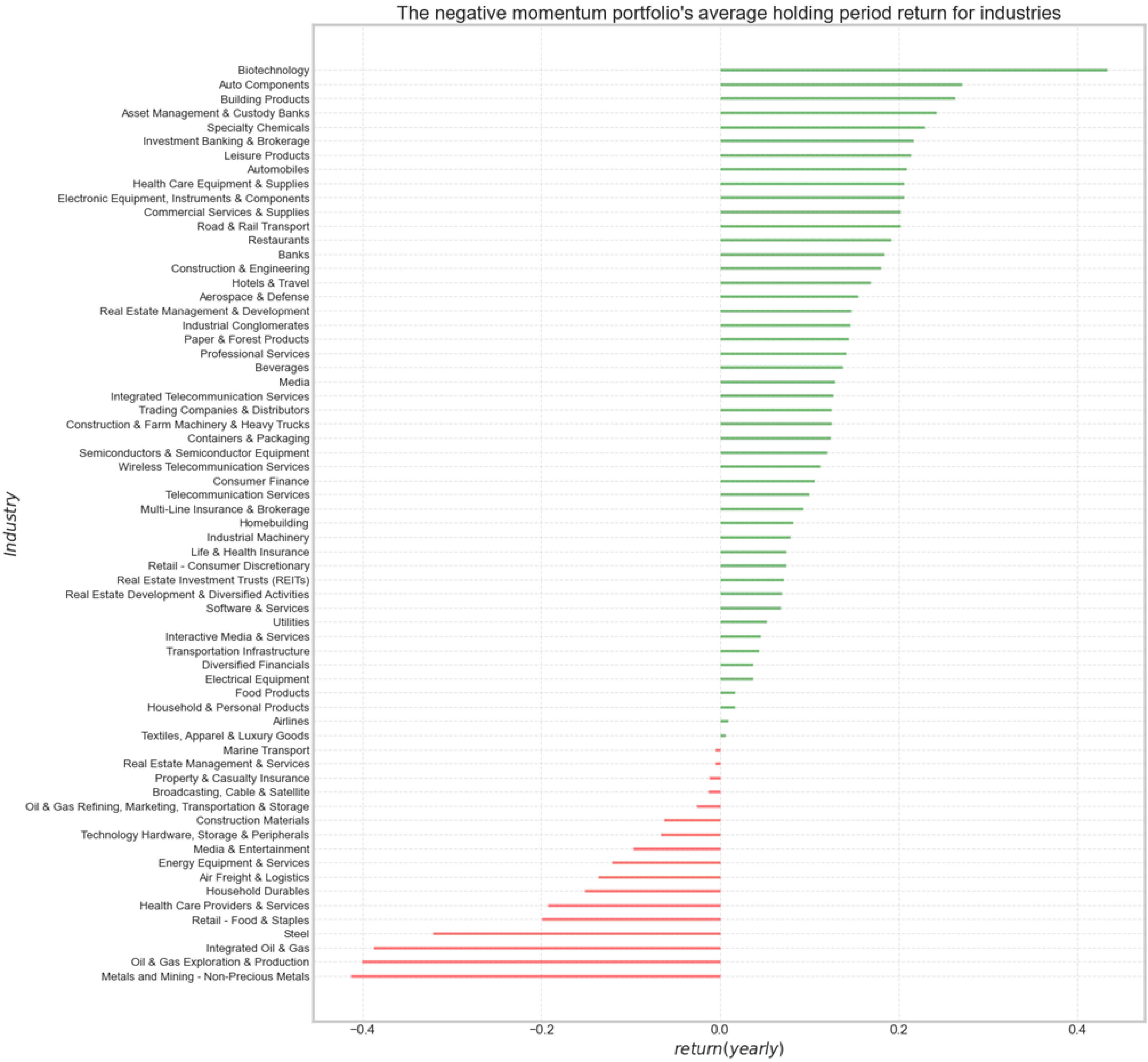

Figure 26: Negative momentum average HPR for industries, using $1\lambda_Y$

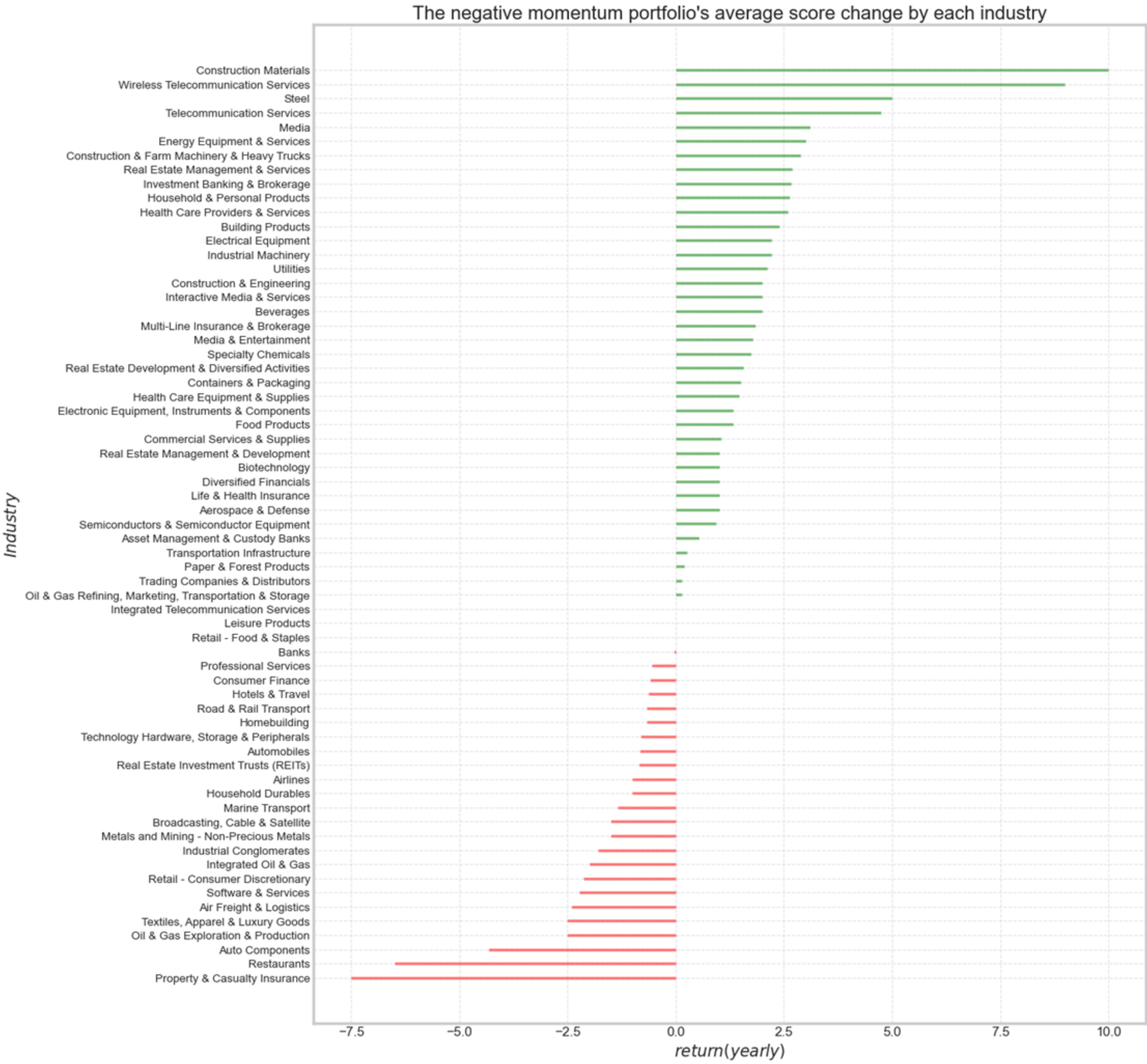

Figure 27: Negative momentum average score change for industries, using $1\lambda_Y$

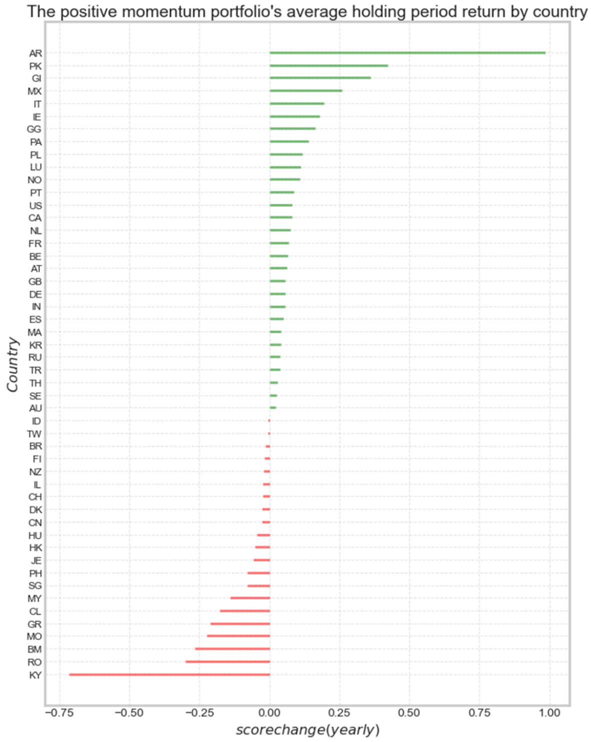

Figure 28: Positive momentum average HPR for countries, using $1\lambda_Y$

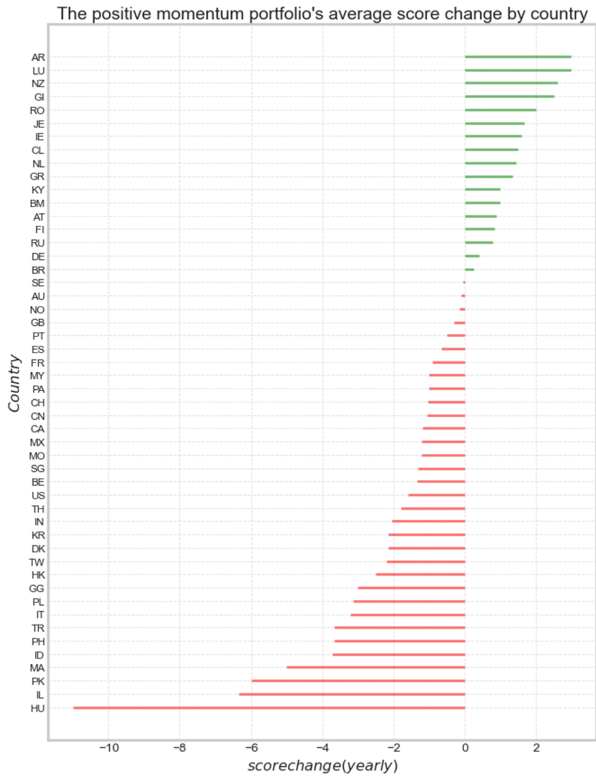

Figure 29: Positive momentum average score change for countries, using $1\lambda_Y$

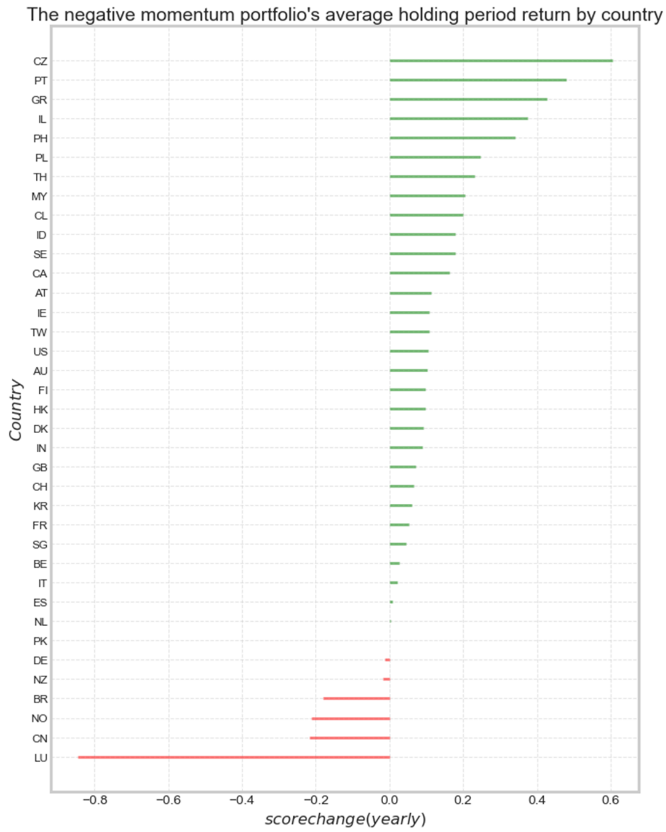

Figure 30: Negative momentum average HPR for countries, using $1\lambda_Y$

Figure 31: Negative momentum average score change for countries, using $1\lambda_Y$

		Coefficient	P-value
portfolio	**Variable**		
Negative	const	-0.001230	0.5713
	WORLD ESG LEADERS Standard (Large+Mid Cap)	0.996805	0.0000
	SMB	0.433451	0.1694
	HML	0.271767	0.3542
	const	-0.000728	0.7479
	WORLD ESG LEADERS Standard (Large+Mid Cap)	0.991855	0.0000
	SMB	0.444079	0.1738
	HML	0.503311	0.1975
	RMW	0.351220	0.4746
	CMA	-0.712151	0.2476
Positive	const	-0.004153	0.0471
	WORLD ESG LEADERS Standard (Large+Mid Cap)	1.024929	0.0000
	SMB	0.643179	0.0336
	HML	-0.187732	0.5010
	const	-0.003863	0.0724
	WORLD ESG LEADERS Standard (Large+Mid Cap)	1.026811	0.0000
	SMB	0.598906	0.0523
	HML	0.247518	0.4985
	RMW	0.164658	0.7211
	CMA	-1.146706	0.0493

Figure 32: Factor analysis for $1\lambda_Y$

D Trend model tables

D.1 Out of sample full set

	Correct	B1 correct	B2 correct	B3 correct	Wrong	B1 wrong	B2 wrong	B3 wrong	No Guess	B1 No Guess	B2 No Guess	B3 No Guess	Guesses correct	B1 guesses correct	B2 guesses correct	B3 guesses correct
0	0.22	0.37	0.26	0.21	0.14	0.33	0.44	0.25	0.63	0.30	0.30	0.54	0.60	0.53	0.37	0.45

D.2 P1: Most guesses correct

	Correct	B1 correct	B2 correct	B3 correct	Wrong	B1 wrong	B2 wrong	B3 wrong	No Guess	B1 No Guess	B2 No Guess	B3 No Guess	Guesses correct	B1 guesses correct	B2 guesses correct	B3 guesses correct
0	0.22	0.37	0.26	0.22	0.14	0.33	0.44	0.26	0.64	0.29	0.29	0.53	0.60	0.53	0.37	0.45

D.3 P2: The least amount of squared errors

	Correct	B1 correct	B2 correct	B3 correct	Wrong	B1 wrong	B2 wrong	B3 wrong	No Guess	B1 No Guess	B2 No Guess	B3 No Guess	Guesses correct	B1 guesses correct	B2 guesses correct	B3 guesses correct
0	0.22	0.37	0.26	0.21	0.15	0.33	0.44	0.26	0.63	0.30	0.30	0.53	0.59	0.53	0.37	0.45

D.4 P3: Most Correct

	Correct	B1 correct	B2 correct	B3 correct	Wrong	B1 wrong	B2 wrong	B3 wrong	No Guess	B1 No Guess	B2 No Guess	B3 No Guess	Guesses correct	B1 guesses correct	B2 guesses correct	B3 guesses correct
0	0.22	0.38	0.26	0.21	0.14	0.33	0.45	0.26	0.64	0.29	0.29	0.53	0.61	0.53	0.36	0.44

D.5 P4: The least amount of wrongs

	Correct	B1 correct	B2 correct	B3 correct	Wrong	B1 wrong	B2 wrong	B3 wrong	No Guess	B1 No Guess	B2 No Guess	B3 No Guess	Guesses correct	B1 guesses correct	B2 guesses correct	B3 guesses correct
0	0.22	0.37	0.26	0.22	0.14	0.33	0.44	0.26	0.63	0.30	0.30	0.52	0.61	0.52	0.37	0.45

Table 14: Momentum trend model's training results

	window	w_lt	lag	strictness	s_lt	maxError	No Guess	Guesses correct	B1 guesses correct	B2 guesses correct	B3 guesses correct
0	8	12	1	0.02	0.02	5	0.74	0.57	0.50	0.41	0.47
0	8	14	1	0.02	0.02	5	0.77	0.57	0.50	0.41	0.46
0	10	12	1	0.02	0.02	5	0.69	0.61	0.50	0.39	0.45
0	10	14	1	0.02	0.02	5	0.73	0.62	0.51	0.39	0.46
0	12	12	1	0.02	0.02	5	0.66	0.52	0.50	0.38	0.47
0	12	14	1	0.02	0.02	5	0.70	0.57	0.51	0.37	0.41
0	8	12	1	0.02	0.05	5	0.75	0.57	0.50	0.41	0.45
0	8	14	1	0.02	0.05	5	0.77	0.55	0.49	0.42	0.46
0	10	12	1	0.02	0.05	5	0.70	0.60	0.50	0.39	0.46
0	10	14	1	0.02	0.05	5	0.73	0.61	0.51	0.39	0.45
0	12	12	1	0.02	0.05	5	0.68	0.53	0.50	0.38	0.43
0	12	14	1	0.02	0.05	5	0.70	0.56	0.51	0.38	0.48
0	8	12	1	0.05	0.02	5	0.74	0.56	0.50	0.41	0.44
0	8	14	1	0.05	0.02	5	0.78	0.55	0.49	0.42	0.42
0	10	12	1	0.05	0.02	5	0.69	0.61	0.50	0.39	0.47
0	10	14	1	0.05	0.02	5	0.73	0.62	0.51	0.39	0.42
0	12	12	1	0.05	0.02	5	0.68	0.53	0.50	0.38	0.47
0	12	14	1	0.05	0.02	5	0.70	0.57	0.51	0.37	0.41
0	8	12	1	0.05	0.05	5	0.75	0.55	0.50	0.41	0.40
0	8	14	1	0.05	0.05	5	0.78	0.53	0.49	0.42	0.47
0	10	12	1	0.05	0.05	5	0.70	0.60	0.50	0.39	0.47
0	10	14	1	0.05	0.05	5	0.74	0.61	0.51	0.39	0.46
0	12	12	1	0.05	0.05	5	0.68	0.53	0.50	0.38	0.43
0	12	14	1	0.05	0.05	5	0.71	0.56	0.51	0.38	0.49
0	8	12	1	0.08	0.02	5	0.75	0.56	0.50	0.41	0.44
0	8	14	1	0.08	0.02	5	0.78	0.55	0.50	0.41	0.43
0	10	12	1	0.08	0.02	5	0.70	0.61	0.51	0.39	0.44
0	10	14	1	0.08	0.02	5	0.74	0.62	0.50	0.39	0.43
0	12	12	1	0.08	0.02	5	0.69	0.52	0.50	0.38	0.45
0	12	14	1	0.08	0.02	5	0.70	0.55	0.51	0.38	0.44
0	8	12	1	0.08	0.05	5	0.76	0.54	0.50	0.41	0.47
0	8	14	1	0.08	0.05	5	0.79	0.54	0.50	0.41	0.48
0	10	12	1	0.08	0.05	5	0.71	0.60	0.51	0.39	0.45
0	10	14	1	0.08	0.05	5	0.74	0.61	0.51	0.39	0.44
0	12	12	1	0.08	0.05	5	0.69	0.52	0.50	0.38	0.46
0	12	14	1	0.08	0.05	5	0.71	0.56	0.51	0.38	0.46
0	8	12	1	0.02	0.02	10	0.71	0.57	0.50	0.41	0.47
0	8	14	1	0.02	0.02	10	0.75	0.56	0.50	0.41	0.42
0	10	12	1	0.02	0.02	10	0.66	0.61	0.51	0.39	0.42
0	10	14	1	0.02	0.02	10	0.71	0.60	0.50	0.40	0.45
0	12	12	1	0.02	0.02	10	0.59	0.56	0.51	0.38	0.45
0	12	14	1	0.02	0.02	10	0.65	0.62	0.51	0.38	0.46
0	8	12	1	0.02	0.05	10	0.72	0.56	0.50	0.41	0.46
0	8	14	1	0.02	0.05	10	0.75	0.54	0.50	0.41	0.49
0	10	12	1	0.02	0.05	10	0.67	0.61	0.51	0.39	0.47
0	10	14	1	0.02	0.05	10	0.71	0.59	0.50	0.40	0.51
0	12	12	1	0.02	0.05	10	0.61	0.57	0.51	0.38	0.43
0	12	14	1	0.02	0.05	10	0.66	0.62	0.51	0.38	0.43
0	8	12	1	0.05	0.02	10	0.72	0.56	0.50	0.41	0.47
0	8	14	1	0.05	0.02	10	0.75	0.53	0.50	0.41	0.45
0	10	12	1	0.05	0.02	10	0.67	0.62	0.50	0.39	0.45
0	10	14	1	0.05	0.02	10	0.72	0.60	0.50	0.40	0.42

0	8	14	1	0.08	0.02	10	0.75	0.54	0.50	0.41	0.44
0	10	12	1	0.08	0.02	10	0.68	0.61	0.50	0.39	0.44
0	10	14	1	0.08	0.02	10	0.72	0.60	0.50	0.40	0.52
0	12	12	1	0.08	0.02	10	0.63	0.57	0.51	0.38	0.45
0	12	14	1	0.08	0.02	10	0.67	0.60	0.51	0.38	0.47
0	8	12	1	0.08	0.05	10	0.73	0.55	0.50	0.41	0.43
0	8	14	1	0.08	0.05	10	0.76	0.53	0.50	0.41	0.46
0	10	12	1	0.08	0.05	10	0.68	0.60	0.50	0.39	0.45
0	10	14	1	0.08	0.05	10	0.72	0.59	0.50	0.40	0.47
0	12	12	1	0.08	0.05	10	0.63	0.57	0.51	0.38	0.48
0	12	14	1	0.08	0.05	10	0.67	0.61	0.51	0.37	0.46
0	8	12	1	0.02	0.02	20	0.69	0.56	0.49	0.41	0.48
0	8	14	1	0.02	0.02	20	0.74	0.57	0.50	0.41	0.49
0	10	12	1	0.02	0.02	20	0.62	0.60	0.50	0.40	0.43
0	10	14	1	0.02	0.02	20	0.69	0.58	0.50	0.40	0.44
0	12	12	1	0.02	0.02	20	0.55	0.58	0.51	0.37	0.42
0	12	14	1	0.02	0.02	20	0.63	0.60	0.51	0.37	0.40
0	8	12	1	0.02	0.05	20	0.70	0.56	0.49	0.41	0.45
0	8	14	1	0.02	0.05	20	0.74	0.55	0.49	0.42	0.43
0	10	12	1	0.02	0.05	20	0.62	0.60	0.50	0.40	0.44
0	10	14	1	0.02	0.05	20	0.69	0.56	0.50	0.40	0.47
0	12	12	1	0.02	0.05	20	0.57	0.58	0.51	0.37	0.44
0	12	14	1	0.02	0.05	20	0.63	0.60	0.51	0.37	0.43
0	8	12	1	0.05	0.02	20	0.70	0.56	0.49	0.41	0.44
0	8	14	1	0.05	0.02	20	0.74	0.55	0.50	0.41	0.48
0	10	12	1	0.05	0.02	20	0.62	0.61	0.50	0.40	0.47
0	10	14	1	0.05	0.02	20	0.69	0.59	0.50	0.40	0.44
0	12	12	1	0.05	0.02	20	0.57	0.58	0.51	0.37	0.44
0	12	14	1	0.05	0.02	20	0.63	0.60	0.51	0.37	0.40
0	8	12	1	0.05	0.05	20	0.71	0.55	0.49	0.41	0.45
0	8	14	1	0.05	0.05	20	0.75	0.53	0.50	0.41	0.43
0	10	12	1	0.05	0.05	20	0.63	0.61	0.50	0.40	0.42
0	10	14	1	0.05	0.05	20	0.70	0.58	0.50	0.40	0.47
0	12	12	1	0.05	0.05	20	0.57	0.58	0.51	0.37	0.46
0	12	14	1	0.05	0.05	20	0.64	0.59	0.51	0.37	0.45
0	8	12	1	0.08	0.02	20	0.70	0.55	0.49	0.41	0.42
0	8	14	1	0.08	0.02	20	0.75	0.54	0.50	0.41	0.43
0	10	12	1	0.08	0.02	20	0.63	0.60	0.50	0.40	0.43
0	10	14	1	0.08	0.02	20	0.70	0.58	0.50	0.40	0.40
0	12	12	1	0.08	0.02	20	0.58	0.58	0.51	0.37	0.47
0	12	14	1	0.08	0.02	20	0.64	0.59	0.51	0.37	0.45
0	8	12	1	0.08	0.05	20	0.71	0.55	0.49	0.41	0.47
0	8	14	1	0.08	0.05	20	0.76	0.53	0.50	0.41	0.48
0	10	12	1	0.08	0.05	20	0.64	0.59	0.50	0.40	0.45
0	10	14	1	0.08	0.05	20	0.70	0.56	0.50	0.40	0.48
0	12	12	1	0.08	0.05	20	0.58	0.58	0.51	0.37	0.43
0	12	14	1	0.08	0.05	20	0.64	0.59	0.51	0.37	0.47

The table shows the results from the last training set done with the trend model.

E Code snippets

```python
def correct_gaps(temp):
    temp = temp.sort_values(by='date')  # Ensure it's sorted by date
    temp['date_diff'] = temp['date'].diff()
    temp['gap_flag'] = (temp['date_diff'] > pd.Timedelta(31, 'D')).cumsum() #If gap is bigger than a month
    temp['isin'] = temp.groupby('gap_flag')['isin'].transform(lambda x: x.iloc[0] + '_' + str(x.name)) #if yes, add "_" and a number to isin
    temp = temp.drop(columns=['date_diff', 'gap_flag'])
    return temp

data2 = data.reset_index().groupby('isin').apply(correct_gaps)
data2 = data2.set_index(['isin', 'date'])
```

Figure 33: Code for handling gaps in data

The code iterated through each company based on dates and when a gap is spotted, we assign a new ID to the company to distinguish them

```python
def pooledRegression(df, periods, factor='msciID', freq='Q', pillar='ESG', alpha=0.05, pct=True):
    temp = df.copy()
    for period in periods:
        if pct == True:
            temp[pillar+'change'+str(period)] = temp.groupby([factor])[pillar].pct_change(period)
            #temp['ESGchangeLag'+str(period)] = temp.groupby([factor])['ESGchange'+str(period)].shift(1)
            temp[pillar+'changeLag'+str(period)] = temp.groupby([factor])[pillar+'change'+str(period)].shift(1)

        else:
            temp[pillar+'change'+str(period)] = temp.groupby([factor])[pillar].diff(period)
            temp[pillar+'changeLag'+str(period)] = temp.groupby([factor])[pillar+'change'+str(period)].shift(1)
    temp = temp.dropna()
    temp = temp.replace([np.inf,-np.inf], np.nan) # replace inf with nan
    temp = temp.fillna(method='ffill')            # forwardfill nans converted from inf.

    pooled_y = temp[pillar+'change1']
    if freq == 'M':
        pooled_X = temp[[pillar+'changeLag1',pillar+'changeLag3',pillar+'changeLag6',pillar+'changeLag12']]
    elif freq == 'Q':
        pooled_X = temp[[pillar+'changeLag1',pillar+'changeLag2',pillar+'changeLag3',pillar+'changeLag4']]
    elif freq == 'Y':
        pooled_X = temp[[pillar+'changeLag1',pillar+'changeLag2',pillar+'changeLag3',pillar+'changeLag4',pillar+'changeLag5',pillar+'changeLag6',pillar+'changeLag7']]
    pooled_X = sm.add_constant(pooled_X)

    # Run regression
    pooled_olsr_model = sm.OLS(endog=pooled_y,exog=pooled_X)
    pooled_olsr_model_results = pooled_olsr_model.fit()

    # OLS assumptions
        # Whites-test for heteroskedasticity
    white_test = dg.het_white(pooled_olsr_model_results.resid, pooled_olsr_model_results.model.exog)

        # Breusch-Godfreys test for autocorrelation
    breuschGodfrey_test = dg.acorr_breusch_godfrey(pooled_olsr_model_results,nlags=periods[-1])

    if (white_test[3] <= alpha) & (breuschGodfrey_test[3] <= alpha):
        #pooled_olsr_model_results = pooled_olsr_model_results.get_robustcov_results(cov_type='HAC',maxlags=periods[-1])
        pooled_olsr_model_results = sm.OLS(endog=pooled_y,exog=pooled_X).fit(cov_type='HAC',cov_kwds={'maxlags':periods[-1]})
    elif (white_test[3] <= alpha) & (breuschGodfrey_test[3] > alpha):
        #pooled_olsr_model_results = pooled_olsr_model_results.get_robustcov_results(cov_type='HC1')
        pooled_olsr_model_results = sm.OLS(endog=pooled_y,exog=pooled_X).fit(cov_type='HC1')
    else:
        pooled_olsr_model_results = pooled_olsr_model_results

    pooled_olsr_model_results.summary()
    return pooled_olsr_model_results, white_test[3], breuschGodfrey_test[3], periods[-1]
```

Figure 34: Code for pooled regression

The code creates the different λ's chosen and run the pooled regression for both quarterly and yearly frequency for MSCI and Refinitiv data set, as well as changing to robust standard erros if necessary.

```python
def fixedEffectsRegression(df,periods,freq='Q',pillar='ESG',alpha=0.05,pct=True,ind=True):
    temp = df.copy()
    for period in periods:
        if pct == True:
            temp[pillar+'change'+str(period)] = temp.groupby(['msciID'])[pillar].pct_change(period)
            temp[pillar+'changeLag'+str(period)] = temp.groupby(['msciID'])[pillar+'change'+str(period)].shift(1)
        else:
            temp[pillar+'change'+str(period)] = temp.groupby(['msciID'])[pillar].diff(period)
            temp[pillar+'changeLag'+str(period)] = temp.groupby(['msciID'])[pillar+'change'+str(period)].shift(1)
    temp = temp.dropna()
    temp = temp.replace([np.inf,-np.inf], np.nan) # replace inf with nan
    temp = temp.fillna(method='ffill')            # forwardfill nans converted from inf.

    if ind == True:
        df_dummies = pd.get_dummies(temp['ind'])
        df_withDummies = temp.join(df_dummies)
        df_withDummies.columns = df_withDummies.columns.str.replace('[&]', '_', regex=True)
        df_withDummies.columns = df_withDummies.columns.str.replace('[-,.() `]', '', regex=True)

        unitNames = pd.Series(temp['ind'].drop_duplicates(keep='first').values)
        unitNames = unitNames.replace('[&]', '_', regex=True)
        unitNames = unitNames.replace('[-,.() `]', '', regex=True)
    else:
        df_dummies = pd.get_dummies(temp['country'])
        df_withDummies = temp.join(df_dummies)

        unitNames = pd.Series(temp['country'].drop_duplicates(keep='first').values)

    y = pillar+'change1'
    if freq == 'M':
        X_names = df_withDummies[[pillar+'changeLag1',pillar+'changeLag3',pillar+'changeLag6',pillar+'changeLag12']].columns
    elif freq == 'Q':
        X_names = df_withDummies[[pillar+'changeLag1',pillar+'changeLag2',pillar+'changeLag3',pillar+'changeLag4']].columns
    elif freq == 'Y':
        X_names = df_withDummies[[pillar+'changeLag1',pillar+'changeLag2',pillar+'changeLag3',pillar+'changeLag4',pillar+'changeLag5',pillar+'changeLag6',pillar+'changeLag7']].columns
    lsdv_expr = y + ' ~ '
    i = 0
    for X_name in X_names:
        if i > 0:
            lsdv_expr = lsdv_expr + ' + ' + X_name
        else:
            lsdv_expr = lsdv_expr + X_name
        i = i + 1
    for dummy_name in unitNames[:-1]:
        lsdv_expr = lsdv_expr + ' + ' + dummy_name

    lsdv_model = smf.ols(formula=lsdv_expr, data=df_withDummies)
    lsdv_model_results = lsdv_model.fit()

    # OLS assumptions
        # Whites-test for heteroskedasticity
    white_test = dg.het_white(lsdv_model_results.resid, lsdv_model_results.model.exog)

        # Breusch-Godfreys test for autocorrelation
    breuschGodfrey_test = dg.acorr_breusch_godfrey(lsdv_model_results,nlags=periods[-1])

    if (white_test[3] <= alpha) & (breuschGodfrey_test[3] <= alpha):
        #lsdv_model_results = lsdv_model_results.get_robustcov_results(cov_type='HAC',maxlags=periods[-1])
        lsdv_model_results = smf.ols(formula=lsdv_expr, data=df_withDummies).fit(cov_type='HAC',cov_kwds={'maxlags':periods[-1]})
    elif (white_test[3] <= alpha) & (breuschGodfrey_test[3] > alpha):
        #lsdv_model_results = lsdv_model_results.get_robustcov_results(cov_type='HC1')
        lsdv_model_results = smf.ols(formula=lsdv_expr, data=df_withDummies).fit(cov_type='HC1')
    else:
        lsdv_model_results = lsdv_model_results

    lsdv_model_results.summary()
    return lsdv_model_results, df_withDummies, unitNames, y, X_names
```

Figure 35: Code for fixed-effect regressions

The code the different λ's chosen in addition to the dummy variables for industry and country and run the fixed-effect regression for both quarterly and yearly frequency for MSCI and Refinitiv data set. It does also make sure we use robust standard errors if necessary.

REFERENCES

Bekaert, G., Rothenberg, R. V., & Noguer, M. (2023, May). *Sustainable Investment - Exploring the Linkage between Alpha, ESG, and SDG's* [SSRN Scholarly Paper]. Rochester, NY. Retrieved 2023-06-28, from https://papers.ssrn.com/abstract=3623459 doi: 10.2139/ssrn.3623459

Berg, F., Koelbel, J. F., Pavlova, A., & Rigobon, R. (2022, October). ESG Confusion and Stock Returns:Tackling the Problem of Noise. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3941514

Berg, F., Kölbel, J. F., & Rigobon, R. (2019, August). *Aggregate Confusion: The Divergence of ESG Ratings* [SSRN Scholarly Paper]. Rochester, NY. Retrieved 2023-01-15, from https://papers.ssrn.com/abstract=3438533 doi: 10.2139/ssrn.3438533

Bonham, J., & Riggs-Cragun, A. (2022, December). *Contracting on Value* [SSRN Scholarly Paper]. Rochester, NY. Retrieved 2023-01-15, from https://papers.ssrn.com/abstract=3892838 doi: 10.2139/ssrn.3892838

Brooks, C. (2019). *Introductory Econometrics for Finance* (4th ed.). Cambridge University Press.

Crespi, F., & Migliavacca, M. (2020, January). The Determinants of ESG Rating in the Financial Industry: The Same Old Story or a Different Tale? *Sustainability*, *12*(16), 6398. Retrieved 2023-01-15, from https://www.mdpi.com/2071-1050/12/16/6398 (Number: 16 Publisher: Multidisciplinary Digital Publishing Institute) doi: 10.3390/su12166398

Damodaran, A. (2008, December). *What is the Riskfree Rate? A Search for the Basic Building Block* [SSRN Scholarly Paper]. Rochester, NY. Retrieved 2023-06-26, from https://papers.ssrn.com/abstract=1317436 doi: 10.2139/ssrn.1317436

Dimson, E., Karakaş, O., & Li, X. (2015, December). Active Ownership. *The Review of Financial Studies*, *28*(12), 3225–3268. Retrieved 2023-01-15, from https://doi.org/10.1093/rfs/hhv044 doi: 10.1093/rfs/hhv044

D'Amato, V., D'Ecclesia, R., & Levantesi, S. (2022, June). ESG score prediction through random forest algorithm. *Computational Management Science*, *19*(2), 347–373. Retrieved 2023-01-15, from https://doi.org/10.1007/s10287-021-00419-3 doi: 10.1007/s10287-021-00419-3

Ehlers, T., Elsenhuber, U., Jegarasasingam, A., & Jondeau, E. (2022, March). *Deconstructing ESG scores: How to invest with your own criteria* [SSRN

Scholarly Paper]. Rochester, NY. Retrieved 2023-06-28, from `https://papers.ssrn.com/abstract=4058612` doi: 10.2139/ssrn.4058612

Environmental, Social and Governance (ESG) Scores from Refinitiv. (2020, April).

ESG Investing. (n.d.). Retrieved 2023-06-28, from `https://www.msci.com/our-solutions/esg-investing`

ESG Investing: ESG Ratings. (n.d.). Retrieved 2023-01-15, from `https://www.msci.com/our-solutions/esg-investing/esg-ratings`

European Central Bank. (2022). *Can EU bonds serve as euro-denominated safe assets?.* LU: Publications Office. Retrieved 2023-06-27, from `https://data.europa.eu/doi/10.2866/84563`

Fama, E. F., & French, K. R. (2014, September). *A Five-Factor Asset Pricing Model* [SSRN Scholarly Paper]. Rochester, NY. Retrieved 2023-06-13, from `https://papers.ssrn.com/abstract=2287202` doi: 10.2139/ssrn.2287202

Giese, G., & Lee, L.-E. (2019). weighing the evidence: ESG and equity returns.

GSIA |. (n.d.). Retrieved 2023-01-15, from `http://www.gsi-alliance.org/`

Kenneth R. French - Data Library. (n.d.). Retrieved 2023-06-27, from `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

Landing Delisted — EOD. (n.d.). Retrieved 2023-07-02, from `https://eodhistoricaldata.com/financial-apis/landing-delisted/`

Looking Back on 30 Years of Carbon Taxes in Sweden. (2020, September). Retrieved 2023-01-15, from `https://taxfoundation.org/sweden-carbon-tax-revenue-greenhouse-gas-emissions/` (Section: Business Taxes)

MSCI ESG Leaders Indexes. (n.d.). Retrieved 2023-06-27, from `https://www.msci.com/msci-esg-leaders-indexes`

MSCI ESG Ratings Methodology. (2022).

Organization for Economic Co-operation and Development. (1960, January). *Interest Rates: 3-Month or 90-Day Rates and Yields: Interbank Rates: Total for Germany.* Retrieved 2023-06-27, from `https://fred.stlouisfed.org/series/IR3TIB01DEM156N` (Publisher: FRED, Federal Reserve Bank of St. Louis)

Organization for Economic Co-operation and Development. (1964, June). *3-Month or 90-day Rates and Yields: Interbank Rates for the United States.* Retrieved 2023-06-27, from https://fred.stlouisfed.org/series/IR3TIB01USM156N (Publisher: FRED, Federal Reserve Bank of St. Louis)

Sankar, A., Lucet, L., Agarwal, N., Haddad, V., & Ouaknine, Y. (2019, March). *ESG Rating and Momentum.* Retrieved 2023-06-28, from https://corpgov.law.harvard.edu/2019/03/21/esg-rating-and-momentum/

Shanaev, S., & Ghimire, B. (2022, May). When ESG meets AAA: The effect of ESG rating changes on stock returns. *Finance Research Letters*, *46*, 102302. Retrieved 2023-01-15, from https://www.sciencedirect.com/science/article/pii/S1544612321003342 doi: 10.1016/j.frl.2021.102302

Smith, K. G., Ferrier, W. J., & Ndofor, H. (2005). Competitive Dynamics Research. In *The Blackwell Handbook of Strategic Management* (pp. 309–354). John Wiley & Sons, Ltd. Retrieved 2023-06-01, from https://onlinelibrary.wiley.com/doi/abs/10.1111/b.9780631218616.2006.00012.x (Section: 11 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/b.9780631218616.2006.00012.x) doi: 10.1111/b.9780631218616.2006.00012.x

Support, E. H. D. (2021, April). *End-Of-Day Historical Stock Market Data API — EOD.* Retrieved 2023-07-02, from https://eodhistoricaldata.com/financial-apis/api-for-historical-data-and-volumes/

What Is Survivorship Bias? Definition and Use in Investing. (n.d.). Retrieved 2023-06-10, from https://www.investopedia.com/terms/s/survivorshipbias.asp

Yahya, H. B., & Vaihekoski, M. (2021). ESG and firm performance: Analysis from Nordic countries' perspective. Retrieved from https://www.efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2022-Rome/papers/EFMA%202022_stage-3032_question-Full%20Paper_id-514.pdf

Zoppe, A., & Lenzi, S. (n.d.). Credit Rating for Euro Area Member States and European supranational institutions.