A Work Project, presented as part of the requirements for the Award of a Master's degree in Finance from the Nova School of Business and Economics.

Cross-Asset Strategy: Nordic ESG Equities, Commodity Pairs Trading and FX Insights

Bård Andre Hansen

Ciaran Jack Callaghan

Jaqueline Alt

Joel Benjamin Lidén

Shiony Wolko

Work project carried out under the supervision of:

Nicholas H. Hirschey

20-12-2023

Abstract

This study examines multi-asset portfolio construction using commodities, foreign exchange, and three equity ESG strategies in the Nordic market. Different allocation methods, including the Maximum Sharpe Ratio and asset weighted strategies, were analysed to understand how non-financial factors such as ESG and Commodity Terms of Trade Signals can impact portfolio performance. The Asset-Weighted strategy, with its strategic diversification across multiple asset classes, delivered consistently high returns and maintained elevated Sharpe Ratios, adeptly balancing risk and return in line with the shifting dynamics of the market.

Keywords: *Finance, Financial Markets, Data Analysis, Equity, Factor-Based Investing, ESG, Commodity Terms of Trade, Foreign Exchange, Currency, Commodities, Systematic Trading Strategy, Currency Pairs, Long Short, Sustainability, Sustainable Investment, Cointegration, , Futures, Quantitative Strategies, Pairs Trading, ESG Momentum, Nordic Stock Market, Sustainable Investment, ESG Rating, Machine Learning, Gradient Boosting Tree, SHAP*

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209)

1 Introduction

Constructing a multi-asset portfolio can be a valuable method of building a successful strategy, providing benefits from diversification and a balanced approach to managing risk and return. The key to this approach is leveraging the strengths of different asset classes to improve portfolio performance and reduce risk. This paper delves into the complexities of constructing a multi-asset portfolio, particularly evaluating various investment strategies.

Firstly, a literature review and economic rationale of individual strategies are provided, including an overview of Environmental, Social, and Governance investing (ESG), equity investment strategies, and strategies focusing on commodities and FX.

After providing a comprehensive explanation of each tactic, the techniques and fundamental principles of each approach are elucidated. Including ESG criteria in various strategies conforms to the rising inclination towards ethical investment and underscores its mounting significance in contemporary portfolio management.

To evaluate the effectiveness of the combined strategies in a multi-asset portfolio context, a detailed analysis of the portfolios' performance and characteristics is provided in the subsequent chapters. This assessment pertains to both the in-sample and out-of-sample performance of the portfolios.

To maximise investment returns, portfolio managers often turn to innovative and multi-faceted approaches. This study investigates the potential of combining five different investment strategies - Gradient Boosting Tree, ESG Momentum, Best-in-Class ESG, Dynamic Commodity Pairs and Commodity Terms of Trade FX - to improve portfolio performance. This research aims to investigate whether the combination of these individual strategies can outperform the individual strategies and determine the most effective allocation between them.

Equity investments are investments in companies and offer the potential for high returns, although with relatively high exposure to systematic risk. Company-specific factors and broader economic trends influence their performance.

Foreign exchange trading involves global currencies and is influenced by international economic factors, interest rates and geopolitical events. The inclusion of foreign exchange in an equity-oriented portfolio can offer diversification benefits due to the different market dynamics that influence these asset classes (Kroencke et al., 2011).

The value of commodities such as precious metals, oil and agricultural products are primarily determined by supply and demand dynamics, geopolitical factors and inflation trends (Ng & Pirrong, 1994) Investments in commodities can serve as a hedge against inflation and are often inversely related to equities (Gorton and Rouwenhorst, 2004).

The focus of this study is to examine potential methods of combining traditional equity investments with alternative assets, specifically, foreign exchange (FX) and commodities. This approach aims to broaden the scope of portfolio diversification and risk management beyond traditional equity and fixed-income portfolios. The weighting methodologies used in this study include risk-parity, asset-weighted, maximising Sharpe ratio, and equal-weighted strategies.

Overall, the aim of analysing various allocation strategies is to understand how different combinations and weightings can impact the risk-return characteristics of a multi-asset portfolio.

2 Literature Review and Economic Reasoning of the Individual Strategies

This section provides an in-depth analysis of five distinct investment strategies: commodities, foreign exchange, and three unique approaches to ESG strategies in the Nordic market. These ESG approaches include Momentum, Best-in-Class, and Gradient Boosting Tree (GBT). The discussion will cover the economic rationale behind each strategy, detailing the construction of the respective models and evaluating their performance. Initially, the economic motivations for the three ESG strategies will be addressed from a unified perspective, followed by an individual examination of each strategy.

2.1 ESG Overview

In the field of sustainable investing, the integration and examination of Environmental, Social, and Governance factors play a central role. First recognised in the 2006 United Nations Principles for Responsible Investing, ESG elements are integral to discussions surrounding their impact on achieving sustainable development goals and influencing financial performance.

Extensive financial research has been undertaken to unravel the complexities inherent in ESG metrics, yielding results that indicate an encouraging upward trajectory in their effectiveness and impact. Research such as that of Friede et al. (2015), who conducted a meta-analysis of over 2000 studies, revealed a positive correlation between high ESG ratings and superior financial performance, suggesting that companies with strong ESG practices often outperform those with weaker commitments. Further studies such as NYU Sterns (2020) also establish the connection between ESG and financial performance, ESG performance is positively correlated with various metrics such as operational efficiency, stock performance, and lower cost of capital. Significantly, the analysis underscored that the financial benefits of ESG become more apparent over longer time horizons. In this study, long-term focused research was 76% more likely to observe positive or neutral results, reinforcing the notion that the impact of ESG on financial performance may be more profound and evident over extended periods.

Whether named corporate social responsibility or ESG, the wider concept has demonstrated its staying power and significance in company analysis, indicating its critical role in shaping future business evaluations. Three distinct ESG strategies were developed to harness this evolving trend: a momentum-based approach, a best-in-class method, and an innovative machine learning strategy utilising Gradient Boosting Trees. Each of these strategies will be thoroughly examined in the following sections, providing detailed insights and justifications, reflecting their contributions to the field of ESG investing.

2.2 Gradient Boosting Tree ESG Strategy

As the Environmental, Social, and Governance (ESG) domain expands, the variability in ESG scoring methods has created a murky landscape for investors. A poignant example of this can be seen in the case highlighted by ESG Clarity (2020), where divergent ratings led to the exclusion of Tesla from the S&P 500 ESG Index while retaining companies such as ExxonMobil, notoriously known for their bad ESG practices. This decision, driven by different methodologies and ratings scales employed by ESG agencies, underscores the challenge of achieving a consistent approach to ESG assessment (Hallez, 2022).

Such ambiguity between ESG methodologies and their applications, although it poses a challenge, also underscores ESG's importance in investment decisions. One of the main challenges in ESG investing is this very ambiguity in ESG scoring, where different agencies provide vastly different scores for the same companies, often based on opaque models (Esma, 2021). The lack of transparency in these scores creates a significant obstacle for investors, complicating the process of determining the most accurate and reliable ESG assessments. With over 600 agencies employing ESG scores, the landscape is further muddled, presenting an opportunity to create a model that addresses these inconsistencies.

Integrating machine learning into modelling ESG metrics on stock performance offers a powerful solution to these challenges. Machine learning's ability to analyse large and complex datasets makes it an ideal tool for revealing the intricate relationships between ESG practices and financial performance. This is particularly relevant in the Nordic markets, where companies are often at the forefront of ESG innovation. The advanced ESG practices in this region provide a wealth of diverse data, ideal for machine learning algorithms to analyse and extract meaningful insights.

Furthermore, applying machine learning in this context aligns with economic reasoning. The Nordic market, recognized for its high ESG standards, presents a unique case where the integration of ESG is closely linked to a company’s financial health. Machine learning models can detect patterns that demonstrate how ESG metrics influence operational efficiency, risk mitigation, and long-term profitability, especially in these markets.

Modelling ESG metrics on financial performance using machine learning is not just a strategy but a response to the need for more transparency and accuracy in ESG investing. This approach is particularly beneficial in the Nordic context, where ESG leadership is synonymous with innovation and sustainable financial performance. It allows investors to navigate through the inconsistency of ESG ratings and directly assess the impact of ESG practices on a company's financial health and return potential.

2.3 ESG Momentum Strategy

The rationale for ESG momentum is rooted in the idea that companies with improving ESG scores are increasingly seen as a predictor of positive future stock performance. Similar to stock momentum, the ESG momentum strategy is predicated on the belief that firms with historical improvements in their ESG profiles are likely to experience positive future stock returns, while those failing to address ESG concerns may face declining stock performance.

There are several studies investigating the relationship between ESG scores and corporate financial performance (Albuquerque et al., 2019; Galema et al., 2008; Kempf and Osthoff, 2007; Von Wallis and Klein, 2015). However, only a few studies focus on the changes in the ESG ratings in relation to financial improvement. A critical aspect of ESG momentum research is highlighted in Nagy et al.'s studies (2013; 2016), which contrast various approaches to ESG investing, including the impact of ESG ratings changes on market performance in the US market. Their findings indicate that ESG momentum strategies tend to deliver better risk-adjusted performance and active returns compared to strategies based solely on high ESG scores. Ouaknine et al. (2010) found similar results for the European stock market, where ESG momentum outperformed the positive screening ESG strategy by over 14%.

Extensive research, notably by Khan, Serafeim, and Yoon (2016), found a significant positive relationship between improvements in ESG scores and future stock performance. This relationship indicates that companies with strong and improving ESG practices often exhibit better financial health, making them attractive investment targets. While Kaiser (2020) reveals that improvements in ESG ratings can predict future stock performance, this relationship shows mixed results. Particularly when focusing on strategies based solely on environmental factors improvements generally correlated with negative stock returns. On the other hand, Nagy et al. (2016) demonstrate that trading on positive ESG momentum yielded significant alpha during 2007 and 2015 with an annual excess return of 2.2%. Interestingly, the study found that the majority of these excess returns can be attributed to idiosyncratic risk, which might be related to ESG signals. This implies that the market may not fully price in the nuances of ESG improvements, creating opportunities for informed investors. (Nagy et al., 2016). This lag of markets not yet fully integrating ESG information into stock prices creates opportunities for investors to benefit from the market's slow response to ESG-related improvements (Gloßner, 2017). With its particular emphasis on sustainability, the Nordic market presents a fertile ground for exploiting these market inefficiencies. This trend is supported by evidence that early adopters of corporate policies focusing on environmental and social issues often achieve higher stock returns (Eccles et al., 2014). Similarly, Dimson et al. (2015) found that successful shareholder engagement in ESG issues leading to changes in business practices resulted in higher abnormal returns over one year.

While ESG momentum has been less extensively studied compared to traditional ESG strategies, the available research, including works by Berg et al. (2022) and Galema and Gerritsen (2022), indicates its potential to leverage the financial gains associated with continuous improvements in ESG scores over mid to long-term periods.

2.4 Best-in-Class ESG Strategy

Various studies have examined methodologies similar to the best-in-class approach for ESG investing in different global regions. These methodologies may be labelled differently, with terms like best-in-sector also being used. Despite their differences in aspects, such as the proportion of stocks selected for investment, these approaches share a common focus on prioritising the best ESG performers. This paper specifically references influential studies in the field of ESG investment strategies that are frequently cited to ensure a strong analytical basis.

Research has suggested that companies with strong ESG performance may outperform their peers over the long term. This is illustrated by Bennani, Guenedal, and Lepetit's findings that between 2014 and 2017, best-in-class stocks in the Eurozone achieved an annualised excess return of 6.6% compared to their worst-in-class counterparts (Bennani et al., 2018).

Investing in companies with strong ESG performance can help mitigate risks such as reputational, regulatory, and operational risks, which can contribute to more stable and resilient investment portfolios. This is a crucial factor in the increasing focus on ESG integration in sustainable investing strategies, as Eurosif's 2018 report emphasised.

The Best-in-Class ESG Strategy is also key to driving ESG improvements across all industries. By directing investment towards companies that excel in ESG factors, it incentivises other companies to improve their ESG performance, promoting positive changes in corporate behaviour and industry standards (Eccles & Viviers, 2011)

The study by Kempf and Osthoff (2007) is a remarkable example of the effectiveness of the best-in-class approach in the area of socially responsible investing. They developed a trading strategy that involved buying stocks with high socially responsible ratings and selling stocks with low ratings. This method resulted in significant abnormal returns, reaching up to 8.7% per year, highlighting the potential financial benefits of this investment strategy.

It is essential to acknowledge that there is no consistency in the findings of studies on this subject. For instance, Statman and Glushkov (2008) conducted a study from 1992 to 2007, which arrived at a varied conclusion. Although portfolios that follow socially responsible investing initially displayed an advantage in returns, the particular approach used, such as excluding certain stocks from undesirable companies, could potentially nullify these benefits. This outcome supports using the best-in-class technique, which concentrates on socially responsible criteria without eliminating so-called "sinful" companies (Statman and Glushkov, 2008).

In the best-in-class approach, all sectors offered by Refinitiv are considered for investment but are exclusively based on companies with high ESG ratings. The strategy involves taking only long positions and not excluding any sector.

2.5 Commodity Pairs Strategy

The incorporation of commodities into an investment portfolio is rooted in their distinct behaviour relative to traditional asset classes, such as stocks and bonds. Commodities, characterised by tangible and finite resources, not only exhibit low correlation with equities but also lack significant correlation among themselves (Erb and Harvey, 2006). This inherent diversification potential provides investors with a means to mitigate overall portfolio risk, aligning with the fundamental principles of modern portfolio theory as advocated by Harry Markowitz (Markowitz, 1952).

Moreover, commodities offer unique benefits, particularly during inflationary periods, showcasing a positive historical relationship with rising prices (Gorton and Rouwenhorst, 2004). Including commodities in a diversified portfolio serves as a safeguard against the erosive effects of inflation on the real value of investments. Tang and Xiong's research (2012) further bolsters the economic rationale for incorporating commodities by emphasising their role in enhancing long-term risk-adjusted returns. Their findings indicate that commodities contribute positively to portfolio performance over extended time horizons.

The mean-reverting nature inherent in commodities, as found by Bessembinder et al. (1995), suggests a tendency to revert to historical averages or equilibrium levels over time. When commodity prices deviate significantly from their historical averages due to external shocks or price pressure, an opportunity exists to take positions anticipating a mean reversion. This leads to a situation where being both long and short in commodity futures could prove beneficial. This approach aligns with the broader concept of statistical arbitrage, where investors seek to profit from temporary price dislocations.

Furthermore, economic motivation extends beyond risk mitigation, inflation hedging and long-short strategy potential. Commodities, positioned as an asset class, provide exposure to global economic cycles and serve as a barometer of economic health (Erten and Ocampo, 2013). The associated diversification benefits, coupled with the potential for capital appreciation during economic expansions, underscore commodities as a valuable addition to a diversified portfolio.

2.6 Commodity Terms of Trade FX Strategy

The concept of using macroeconomic variables to predict changes in asset prices is far from new. Many of the most common currency trading strategies, such as the carry-trade strategy, the dollar-carry trade strategy and the momentum strategies, attempt to exploit exchange rate movements which deviate from the uncovered interest rate parity condition, which suggests that the difference in interest rates between two countries should be equal to the expected percentage change in the exchange rate between those two currencies. Bilson (1987), Fama (1984), and Froot and Thaler (1990) have all found empirical evidence for deviations from the condition, which imply that interest rates are not the optimal predictor of future exchange rates (Filippou & Taylor, 2017). These deviations imply that other variables contribute to predicting the future exchange rate.

Multiple prior studies have attempted to model real exchange rates using a variety of methodologies, Meese and Rogoff (1983) famously found that a range of models exchange rate models based on macroeconomic fundamentals were unable to predict nominal and real exchange rates more accurately than simple random walk model, this finding was reiterated by Cheung, Chinn and Pasual (2005). Much of the empirical evidence which refutes a quantifiable link between exchange rates was conducted on developed economies, however, more recent studies which focus on commodity-exporting, developing economies claim to overturn these findings, such as Kohlscheen, Avalos and Schrimpf (2017), who found (pseudo) out-of-sample predictability of exchange rates using commodity prices in a sample of commodity-exporting countries, or Cashin, Céspedes and Sahay (2004) who found evidence of a long run relationship between real exchange rate and real commodity prices. Indeed, if exchange rates do follow a random walk, then changes to the exchange rates would persist and vary randomly, as opposed to returning to a long-run equilibrium as proposed by purchasing power parity, which posits that exchange rates should move towards the level that the prices of a basket of goods and services in different countries (Terborgh, 1926).

A macroeconomic variable which theoretically could influence exchange rates is commodity terms of trade. In general, a country's terms of trade, the ratio of their export prices to their import prices, can be interpreted as the quantity of import goods an economy can purchase per unit of their export goods. As one might expect, 'commodity' terms of trade follow the same definition but consider only the ratio of commodity export and import prices.

Ceteris paribus, improvement in terms of trade should be a positive sign for an economy (Collier & Goderis, 2012), imports become cheaper compared with exports and demand for the country's exports increases, which in turn increases demand for the country's currency, causing it to appreciate. Given this theoretical link, it is not hard to imagine that accurate and timely terms of trade data could bring a rich vein of insight to the appropriate market participant. The most common way in which terms of trade changes are measured is through commodity terms of trade indices. These indices weigh the effect of price changes in goods and services based on the contribution of those goods and services to the country in question's economy to determine the net effect of a price change on the country's terms of trade. The problem is that official terms of trade figures are released with a lag as trade data is not reported instantaneously, the figures are entirely compiled in hindsight, using only actual prices paid for completed trade. To solve the problem of data lag, market participants have come to focus more on commodity terms of trade indices. These indices consider a basket of relevant primary commodities, while not encompassing all of a country's trade, their prices tend to vary more than those of more processed goods and services and thus tend to have a larger influence on short-term macroeconomic dynamics (Arezki et al., 2012). Additionally, prices of such commodities are available point-in-time from markets. This strategy will build rolling weight CToT indices for the selected 36 currencies following a methodology similar to that of Gruss and Kebhaj (2019), with key differences implemented to develop indices which can predict the effect of commodity price changes on a country's terms of trade, as well as ensuring they are representative of a country's current trade flows, providing real-time data for use as a signal for trading foreign currencies both long and short against the US dollar in a dollar-denominated portfolio. To the best of the author's knowledge, this strategy has not been tested and published on any public forum.

3 Description of the individual strategies

In this chapter, different investment strategies that are suited to particular market conditions and goals are discussed. These strategies vary from utilising advanced machine learning techniques in ESG investing to employing dynamic approaches in the commodity and foreign exchange markets, giving a thorough outline of various investment methodologies.

3.1 Gradient Boosting Tree ESG Strategy

A well-known model in the field of machine learning is XGBoost, a refined version of Gradient Boosting Trees, known for its ability to efficiently process complex and incomplete datasets, a characteristic often found in ESG data. The model's effectiveness is underscored by its predictive accuracy with unseen datasets, especially structured datasets, which has established it as a favoured tool in machine learning competitions at data platforms such as Kaggle, which indicates its suitability for the task.

The effectiveness of XGBoost was empirically tested across various datasets in the foundational paper by Tianqi Chen and Carlos Guestri (2016). Key results include the algorithm's robust performance across different computational settings, its efficient handling of missing data, and the use of advanced techniques like exact greedy and approximate algorithms to determine the best way to split the data at each node of the decision tree. These features have been shown to improve both the speed and accuracy of the model, indicating the suitability for complex predictive tasks such as modelling stock returns based on ESG metrics.

Despite the extensive use of machine learning in various competitions, the direct application of these techniques, including XGBoost, in forecasting stock returns based on ESG metrics, remains a relatively untapped field. Limited studies, such as Bloomberg (2020) point towards the potential of such approaches, where they managed to outperform both the Russel 1000 and S&P 500 in the selected period.

This strategy, therefore, aims to explore this gap, utilizing XGBoost's proven strengths to provide new insights into the relationship between ESG metrics and stock market performance. Further, SHapley Additive exPlanations (SHAP) introduced by Lundberg and Lee (2017), will be utilized in selecting the top 20 performing ESG metrics. SHAP is an approach based on game theory used to explain the actual predictions of the model as it is of great interest to understand what factors are guiding the predictions.

To evaluate the model's precision, a grid search on optimizing the model's mean squared errors (MSE) for the training period outliers was conducted. A grid search optimizes the Hyperparameters of the model and selects the one which would achieve the best results during the training period.

To summarise, this strategy offers a comprehensive approach to understanding and leveraging ESG metrics in investment decisions. By integrating advanced machine learning techniques like XGBoost and SHAP, this strategy goes beyond traditional analysis to capture the complex relationship between ESG metrics and financial returns. It addresses the need for greater clarity and objectivity in ESG investing, Utilizing one of the more prominent markets – the Nordic stock market.

3.2 ESG Momentum Strategy

While price momentum is a stable part of financial markets as established by Jegadeesh and Titman in 2001 and evident across various markets (Moskowitz, Ooi, and Pedersen, 2012; Hartley, 2020), ESG scores are emerging increasingly. The concept of winners minus losers has extended to socially responsible investing, as Vojtko and Padysak (2019) found that ESG scores can be effectively used in various investment strategies, including negative screening and momentum approaches.

The ESG momentum strategy, a novel approach to sustainable investing, diverges from traditional long-term investment strategies by focusing on the dynamic nature of ESG scores. Central to this strategy is the belief that a company's future stock performance is closely tied to changes in its ESG quality. It operates on the insight that improvements in ESG scores signal a company's better capability to mitigate ESG-related risks, which market participants quickly recognize and factor into the company's share price. Unlike strategies that focus on companies with the highest ESG scores, the ESG momentum strategy targets those exhibiting significant positive changes in their ESG performance.

The cornerstone of this strategy is the computation of the ESG score change, calculated as a 12-month differential, reflecting the annual nature of ESG score reporting. This methodology facilitates the identification of corporations demonstrating significant positive shifts in their ESG profiles, aligning with the strategy's core thesis: firms exhibiting substantial ESG improvements are poised to outperform their counterparts.

Papers like Nagy (2016) showed that the execution of this strategy is distinct from any ESG tilt strategy, which leans towards companies with consistently high ESG ratings. The ESG momentum strategy instead prioritizes companies that have demonstrated a notable increase in their ESG ratings over the past year. This focus does not directly aim to raise the overall ESG profile of the portfolio but identifies potential short-term gains from positive ESG developments.

Moreover, Nofsinger and Varma (2013) posited that higher ESG ratings enhance a company's resilience against market downturns, legal challenges, and regulatory fines. Their findings are bolstered by subsequent studies (e.g., Nagy et al., 2013; Verheyden et al., 2016; Pollard et al., 2018), which collectively underscore the profitability potential of the ESG momentum strategy.

3.3 Best-in-Class ESG Strategy

The Best-in-class approach, deeply rooted in ESG investment principles, focuses on selecting the top 10% of companies based on their ESG performance within each OMX Nordic Large Cap Index sector. This methodological choice is underpinned by reducing sector bias, a concern highlighted by Eurosif (2018), which noted the growing importance of ESG integration and best-in-class strategies in sustainable investing. After determining which stocks are eligible for investment, they are then weighted in the portfolio based on their respective ESG scores to ensure that those with a higher ESG performance have a proportionally greater impact on the investment strategy.

In implementing this approach, the aim is to achieve a balanced and diversified portfolio that accurately reflects the leading sustainability practices in the Nordic market. Through this methodical selection and weighting of the best ESG performers from different sectors, the strategy ensures a comprehensive representation of sustainable excellence in the region.

A range of ten sectors, including Industrials, Financials, Consumer Cyclicals, Healthcare, Technology, Real Estate, Basic Materials, Consumer Non-Cyclicals, Energy and Utilities, are considered in this approach.

It is important to acknowledge that certain industries, such as Consumer Non-Cyclicals, Real Estate, Energy, and Utilities, are combined into a single category called "Other." This decision was made because insufficient data is available for these sectors, especially in the first year of data observation. The purpose of this grouping is to ensure that the assessment and selection process is fair and unbiased across all industries, regardless of any limitations in data or differences in ESG performance that may be inherent to specific sectors.

This strategy follows a trading cycle that spans one year, during which ESG scores are used as the primary criterion for decision-making in January. This annual cycle aims to ensure consistency in the assessment of ESG performance and the composition of the portfolio. It is important to note that companies that are newly included in a given year are only considered for inclusion in the portfolio in the following year's assessment.

3.4 Dynamic Commodity Pairs Strategy

This commodity pairs strategy introduces a dynamic cointegration framework, departing from conventional static models with fixed formation and trading periods. The innovation lies in enhancing the strategy's adaptability to changing market conditions. By incorporating a signal with a rolling cointegration window, the approach provides a more nuanced understanding of specific pairs relationships, enabling investors to capitalise on short-term deviations from their historical means. Moreover, this dynamic framework expands the spectrum of tradable pairs, fostering increased diversification effects and, ideally, creating more profitable trading opportunities.

The approach involves purchasing an underpriced commodity future and concurrently selling an overpriced counterpart relative to each other when the z-score of the pairs spread surpasses a predefined threshold. The determination of the number of contracts for each future is guided by a hedge ratio, ensuring the formation of a zero-cost portfolio. Pairs trading offers a compelling avenue to capitalise on profit opportunities arising from market overreactions or underreactions to new information.

The pairs strategy presented in this paper leverages 17 distinct continuous front-end commodity future contracts, all traded on US exchanges, facilitating a pool of 136 tradable pairs. In conjunction with surpassing the z-score threshold, a signal to initiate a position is contingent upon the pairs demonstrating a cointegration relationship over the preceding 200 trading days, a criterion that must persist for the subsequent 5 days, thus limiting pairs cointegrated by mere chance ability to enter the tradable pool.

Cointegration fundamentally encapsulates the idea of mean reversion in asset prices. When two assets exhibit cointegration, it indicates the presence of an equilibrium relationship between them. In pairs trading, regardless of the chosen selection method, the objective is to leverage deviations from this equilibrium within a nonstationary common factor asset-pricing framework (Nicolas Huck and Komivi Afawubo, 2014).

The cointegration approach is specifically adopted due to its demonstrated econometric reliability in identifying comoving securities. In the literature, articulated by Krauss (2016), there exist several different methodologies within pairs trading literature. While each approach has its merits, the cointegration approach stands out for its ability to establish a more robust equilibrium relationship. Empirical studies, exemplified by the work of Gatev et al. (2006) and subsequently reinforced by Do and Faff (2010), emphasise the effectiveness of cointegration tests in identifying pairs with persistent relationships.

The research conducted contributes to the literature by offering a fresh outlook on commodity pairs trading through the application of a dynamic cointegration approach. Going beyond conventional methodologies, this novel approach creates opportunities for pairs trading on commodities even in the absence of apparent fundamental relationships. The introduction of dynamic cointegration aims to capture evolving interdependencies among commodity pairs, enabling adaptability to changing market conditions. Moreover, the inclusion of commodities lacking evident fundamental connections adds an innovative dimension to the study, challenging conventional notions and expanding the potential scope of trading opportunities.

3.5 Commodity Terms of Trade FX Strategy

The commodity terms of trade strategies were tested on a universe of 36 developed and emerging market currencies, including the US dollar as the base currency, against which the remaining 35 currencies were traded. The decision was made to trade foreign currencies against the US dollar to ensure the most liquid foreign exchange market for each currency pair, aside from a few select European countries whose most liquid market is with the Euro but are traded in sufficient volumes with the US dollar that liquidity was not a concern.

The strategy is unique in its approach to currency trading as it uses the commodity terms of trade of each currency union to determine the weight and direction of trades based on the change relative to the change in the commodity terms of trade of the US. The logic behind this signal is that if a foreign country or currency union's terms of trade have improved more than that of the US, or indeed if the US commodity terms of trade have fallen, then there should be an appreciation in the foreign currency relative to the US dollar as more of the foreign currency is demanded, and relatively less of the US dollar is demanded. In this scenario, the strategy took a long position in the foreign currency against the US dollar in anticipation of a subsequent appreciation. In the case of a decrease in the foreign currency unions commodity terms of trade relative to the US dollar, a short position was taken in the foreign currency as by the same logic the US dollar would be expected to appreciate relative to the foreign currency.

The size of the position as well as the weighting given to the trade in the combined portfolio of currency pairs was determined by the magnitude of the deviation relative to the historical mean, with deviations of higher magnitude being allocated greater weight in the overall portfolio. An advantage of using the historical mean is that as the strategy operates for longer, the mean will become more stable and representative of the 'true mean', in theory allowing the strategy to improve in performance as time progresses.

Deviations were measured over the previous four weeks to ensure rapid or short-lived commodity price fluctuations do not lead to positions taken without sufficient economic reason, while weekly rebalancing still allows the strategy to be dynamic enough to capture significant terms of trade dislocations and leverage them to make profitable trades. Statistical techniques were also employed to ensure that excessively large positions in single markets were not taken, which could compromise the diversification effect of holding a larger basket of currency pairs.

As the commodity dependence of different currency unions varies significantly, the ability of commodity terms of trade to predict changes in each union's exchange rate with the US is also likely to vary. To address this issue, the strategy was first tested on an in-sample testing period before excluding the currency pairs, which did not contribute to the risk-adjusted return of the strategy.

4 Data and Research Methodology

In this chapter, we delineate the methodologies adopted for data sourcing and preparation across the spectrum of strategies under review. Ensuring analytical coherence, particularly in the combined strategy analysis, necessitates a standardisation process wherein the returns from all strategies have been methodically converted to a daily return format. All equity strategies use the OMX Nordic Large Cap, further referred to as OMXNLC, as a benchmark.

4.1 Data and Research Methodology of the individual strategies

4.1.1 Gradient Boosting Tree ESG Strategy

This section describes the dataset selected from the Refinitiv Platform for the years 2007 to 2022, focusing on the OMXNLC stocks return and respective ESG metrics. The periods are divided into training (2007-2013), in-sample(2014-2017), and out-sample(2018-2022) periods. The predictions from 2013 onwards use rolling windows, meaning that the model trains extensively on all available previous data. This is incorporated to make sure that all available data is utilized and helps in back-testing the model. The Nordic market was chosen for its leadership in ESG reporting and the dataset initially contained 729 ESG metrics, reduced to 353 after excluding those with a 100% missing rate or irrelevancy.

The methodology incorporates a Gradient Boosting Tree approach using XGBoost, tailored for handling ESG data complexities like missing values and non-linearity, as outlined in the strategy profile. In the selection of stocks, further exclusion had to be made as OMXNLC in some cases includes A & B stocks, where they are linked to the same ESG metrics. The final dataset consists of 198 stocks for 2022, while the first dataset for 2007 consisted of 105 different stocks. Stocks are evaluated one year after they are introduced to the market, as there is always a lag in the reporting of ESG metrics. The signal used for trading is created by utilizing XGBoost to model ESG metrics to predict a stock's active return, which measures its performance against a benchmark. This strategy is grounded in the belief that stocks excelling in ESG criteria are likely to surpass their market peers. Therefore, the justification for focusing on Active Return as the dependent variable, rather than mere stock return, is to target stocks that achieve outperformance in their respective sectors, attributing success to strong ESG practices rather than general market movements.

To evaluate the performance of the model, a grid search was conducted in the training period, optimizing the mean squared error for top & bottom 15% performers. As it is of interest to understand which features affect the decision-making in the trees, Shapley Additive exPlanations (SHAP) will be employed, helping to interpret the contributions of each feature to the model's predictions. SHAP, based on Shapley values from cooperative game theory, provides interpretable explanations for the predictions of any machine learning model by quantifying the contribution of each feature to individual predictions. a new model will be trained based on the top 20 features selected based on their SHAP values, where features not showing ambiguity are prioritized. This means that a feature that had a positive effect on top values in one year but a negative effect in the next year is excluded.

4.1.2 ESG Momentum Strategy

The data for the ESG momentum strategy was retrieved from Refinitiv Eikon, ranging from 01/01/2007 to 31/12/2022. The data sample consisted of a historical composite of OMXNLC-listed stocks with daily returns and market capitalization, as well as the corresponding stock-specific ESG scores. For each individual year, the actual composition of the OMXNLC was utilized and then merged to ensure the correct compositions and to avoid survivorship bias.

The methodology for constructing ESG momentum portfolios implemented findings from previous studies highlighting the impact of stock selection thresholds on portfolio performance. The approach was rooted in the findings of Jegadeesh and Titman in 1993, who demonstrated that using stocks only in the top and bottom deciles for a long-short portfolio yielded superior results compared to a lower threshold encompassing all stocks in the sample universe. Similarly, Bird et al. (2017) found that including more stocks in momentum portfolios with a lower threshold negated statistically significant returns.

In this study, three distinct ESG momentum portfolios were constructed using a data set based on the evolution of ESG scores: a long-short portfolio, a long-only portfolio, and a short-only portfolio. Each year, companies were categorized as winners or losers based on their ESG score development. Winners were those with the most positive ESG score development over the year, with losers being those with the worst ESG score development. The portfolios were annually reconstructed based on the change in ESG scores from the previous year. For instance, the 2008 portfolios were based on ESG score changes during 2007–2008, and the selected stocks were held for a year before the portfolio was reconstituted. The positive ESG momentum strategy based solely on equities demonstrating ESG score enhancements, generated the strongest returns during the in-sample analysis, thereby serving as the referencing model for the out-of-sample application.

After careful evaluation of various thresholds for equity inclusion in the ESG momentum portfolios, a strategic decision was made to raise the inclusion threshold by dividing the stock universe into quintiles according to ESG score development. The positive ESG momentum portfolio takes a long position in the top 20% (winners), effectively narrowing down the number of stocks in the portfolios. An essential consideration in this methodology is the company-specific risk, which increases significantly when only the top and/or bottom deciles of stocks are included due to the drastic reduction in the number of companies in the portfolios. By focusing only on the top 20%, the strategy seeks returns at the extreme ends of the stock universe and excludes companies with minor changes in their ESG scores. The portfolio was built by building an equal-weight scheme where the chosen stocks are scaled according to their market capitalization.

4.1.3 Best-in-Class ESG Strategy

ESG investing using the Best-in-Class approach aims to create a portfolio that prioritizes responsible investing through a value-weighted selection of stocks. This method focuses on choosing stocks with ESG scores that rank in the top 10% within their sectors. The strategy is specifically designed for the Nordic region and uses the OMXNLC as its primary source of data. This index is an appropriate choice due to the strategy's geographic emphasis.

Data was collected from 2007 to 2022 using Refinitiv Eikon and its Eikon API, with Python used to ensure efficient and accurate data retrieval. This collection included daily close prices, sectors, and ESG scores for 231 unique instruments. The time frame for analysis was split into two parts: an in-sample period from 2007 to 2017, and an out-of-sample period from 2018 to 2022. This division allowed for a thorough analysis and validation of the investing approach.

Every year, the portfolio is created and adjusted to reflect changes in ESG scores, which are updated in December. However, the ESG score used for portfolio creation is applied in January each year to ensure consistency in the annual investment cycle. It's crucial to understand that newly listed companies are only evaluated for inclusion in the following year's portfolio, allowing for a full year's worth of data to be analysed.

To ensure a fair assessment process, sectors with lower data availability, such as Consumer Non-Cyclicals, Real Estate, Energy, and Utilities, were grouped as "Other." This grouping was necessary as a result of varying data availability levels, especially in the earlier years. Despite the differences in ESG performance or inherent data limitations, the decision to combine these sectors is made to maintain balance and fairness.

This methodology represents a thoughtful and data-driven approach to ESG investing in the Nordic region. By focusing on high-performing ESG stocks within their sectors and adjusting for annual changes in ESG scores, the Best-in-Class portfolio aims to be an investment model for responsible investing.

4.1.4 Dynamic Commodity Pairs Strategy

For the empirical analysis of the strategy, the data consists of liquid commodities with concurrent data accessibility from Refinitiv. The dataset that was compiled spans from the first trading day of 2007 through the conclusion of 2022. This dataset is partitioned into an in-sample period, covering 2007 until the end of 2017, and an out-of-sample period encompassing the remainder of the dataset from 2018 to the end of 2022. The data was imported into Python using the Eikon API. The selected commodities were categorised into three main groups: Energy (Crude WTI, Natural Gas, Heating Oil, Gasoline), Metals (Gold, Platinum, Silver), and Agriculture (Wheat, Soybeans, Coffee, Sugar, Corn, Soybean Oil, Lean Hogs, Cotton, Feeder Cattle, Oats). Resulting in the creation of 136 unique pairs possible to trade in the strategy. All the contracts corresponding to the individual commodities were continuous front-end contracts denominated in US dollars.

Cointegration calculations are performed on a 200-day rolling basis for each commodity pair. The choice of a 200-day window was made deliberately, striking a balance between allowing sufficient time for the time series to exhibit dependence and maintaining adaptability to changing market conditions. This window size demonstrated robustness during the in-sample testing phase. Unlike traditional approaches involving a forming period followed by a trading period, cointegration was employed in this paper to function as a trading signal. The Statsmodels Python library's coint() function, which is based on the augmented Engle-Granger two-step cointegration test, is employed as a measure of cointegration in the analysis. This test is designed to evaluate the null hypothesis of no cointegration between variables, assuming that the variables in y0 and y1 are integrated of order 1, I(1).

A pair was considered tradeable if it had been cointegrated for 200 days for at least five consecutive days. This requirement aimed to mitigate sensitivity to signals arising by chance, ensuring that mostly meaningful signals were considered for trading. A synthetic portfolio was formed of each pair, involving the opening of both a long and a short position. In addition to cointegration, the pair's z-score had to deviate from the rolling mean by at least 1.5 standard deviations for a position to be initiated. This strategy effectively involves buying the underpriced future and selling the overpriced one relative to each other. Positions were maintained until the spread surpassed the equilibrium by +/- 0.75 standard deviations. This approach was designed to enable the strategy to capture momentum in the mean-reverting spreads.

To manage unexpected behaviour, positions were closed if a pair did not exhibit the anticipated mean reversion and remained between 1.5 and 2.5 standard deviations from the mean for more than 60 trading days. To safeguard against significant downside risk, the strategy incorporated a parameter whereby positions were automatically closed if the z-score exceeded a threshold of 3.5 or -3.5. This precaution was introduced to account for the volatile nature of the underlying contracts and to prevent the strategy from being adversely affected by temporary large deviations due to external shocks, which did not adhere to the assumptions of the strategy. Lastly, during the in-sample period, the analysis involved a pre-selection phase followed by a post-selection phase. In the post-selection, the strategy discarded the bottom half of pairs from the trading pool based on risk-adjusted returns obtained in the pre-selection phase. This selection criterion is grounded in the rationale that pairs demonstrating ineffectiveness within the strategy framework during the pre-selection period are deemed unsuitable for further trading.

4.1.5 Commodity Terms of Trade Strategy

Given the systematic nature of the strategy, all computations and executions were carried out through a model with predefined parameters in Python. After constructing the CToT indices for each currency, the initial step involved calculating the z-score for each currency's CToT index on every date. To prevent forward-looking bias, only past CToT scores were considered when calculating the mean. Z-scores were then winsorised at two standard deviations to address outliers and mitigate excessive risk-taking in any particular market.

These z-scores played a pivotal role in determining both the weight of each currency in the portfolio and the trade direction (long or short against USD). Currency weights were calculated based on the magnitude of the terms of trade change relative to the USA's terms of trade over the last four weeks. The larger the change, the greater the weight. The trade direction was determined by the sign of the terms of trade change, with a long position taken if the change relative to the USA was positive and vice versa. Normalisation of weights ensured that the sum of their absolute values always equalled one, with weekly rebalancing.

Returns from the strategy encompass two components: the fluctuation in the exchange rate of the currency (exchange rate effect) and the foreign risk-free rate earned in long positions or paid in short positions (interest rate effect). Strategy transaction costs were estimated following Bollerlev et al. (2016), using half the median of the bid/ask spread of each currency over the last nine months. The weighted returns of each foreign currency/USD pair were then aggregated to derive the portfolio return.

The strategy underwent an initial in-sample testing phase from the beginning of 2007 until the end of 2017. Subsequently, all currencies that performed with a negative Sharpe Ratio under the strategy weighting scheme were excluded. The strategy was then assessed during an out-of-sample period, both individually and as part of the combined group portfolio.

4.2 Data and Methodology for the Combined Strategies

The analysis examined a combined investment strategy consisting of five different variations: an equal-weighted approach, an equal-weighted approach based on the asset class, a target volatility approach, a risk-parity approach, and a maximum Sharpe Ratio approach. The investigation covered two distinct periods, namely an in-sample period spanning from 2014 to 2017 and an out-of-sample period spanning from 2018 to 2022. The strategy weights were applied to the returns daily, with the weighting scheme updating annually.

4.2.1 Equal-Weighted Strategy

The equally weighted strategy holds an equally distributed amount of all the individual strategies over the periods. Every strategy is included in the portfolio with a weight of 20% until the end of the sample. This allocation remains unchanged over time, regardless of the fluctuating performance of the individual asset classes.

4.2.2 Asset-weighted Strategy

The asset-weighted portfolio strategy is characterized by its simplicity and diversification approach. This method differs from more traditional weightings, such as equally weighted portfolios, by emphasizing capital allocation based on different asset classes.

In this context, the asset-weighted strategy is applied specifically to three broad asset classes. The first three strategies - Machine Learning Approach, ESG Momentum and Best in Class ESG Strategy - are grouped as one asset class, reflecting their common focus on equity-based investments. The portfolio is then expanded into more asset classes through the inclusion of commodities and foreign exchange, making it a multi-asset portfolio with each class weighted at one-third. Within this framework, the equity investments are equally weighted so that each equity strategy has a weighting of one-sixth.

4.2.3 Risk Parity Strategy

Incorporating insights from the S&P Risk Parity Indices, the risk parity approach for a multi-asset portfolio of equities, commodities, and FX aimed for a balanced risk distribution across these diverse asset classes. A Risk-Parity Portfolio is based on the principle that each asset within the portfolio contributes equally to the overall risk. Unlike traditional portfolios where asset allocation is based on capital, in risk-parity portfolios, allocation is based on the risk contribution of each asset. This means that riskier assets will have a smaller capital allocation, while less risky assets will have a larger capital allocation. The objective is to achieve a balanced risk distribution across various asset classes, leading to more stable and potentially improved risk-adjusted returns over time. (Brznek et al., 2020)

In this strategy, the individual strategies were clustered into the asset classes; equities, commodities, and FX to ensure each contributes equally to the portfolio's overall risk profile, echoing the principles observed in the S&P Risk Parity Indices. The approach requires annual rebalancing, to distribute risk equally and adjust allocations dynamically to maintain a predefined volatility level, thereby aiming for consistent performance across various market conditions.(Brznek et al., 2020)

4.2.4 Maximzed Sharpe Ratio Strategy

The Sharpe Ratio is a measure of the excess return (return above the risk-free rate) per unit of risk in an investment. A maximum Sharpe Ratio strategy seeks to maximize this ratio for the given period. It's a specific application of the mean-variance optimization that not only considers the balance between risk and return but also incorporates the risk-free rate in its calculation. The goal is to find the portfolio that offers the highest excess return for each unit of risk taken. The weighting scheme is optimized in the in-sample period, selecting the returns yielding the best results, where the same weighting is later applied to the out-sample period to test any differences.

5 Performance and characteristics of the portfolios

This section will provide a comprehensive overview of the in-sample performance for each individual strategy, spanning from the inception date of each strategy to the conclusion of 2017. Additionally, it will examine the performance of various combined portfolio strategies employing multiple weighting schemes. The evaluation will cover the period from February 2014, marking the earliest date when all strategies could generate returns, through the end of 2017.

5.1 In-Sample

The table below contains selected summary statistics of each individual strategy, to which the performance descriptions in the following section will refer.

Table 1: *Individual Strategies In-Sample Performance Statistics*

	Return	Volatilit y	Skewne ss	Kurtosis	Sharpe Ratio	Max Drawdown	Transaction Costs
GBT-Model	17.66%	18.60%	-0.11	5.45	0.94	-19.18%	0.00385%
ESG Momentum	5.79%	20.45%	0.09	10.10	0.23	-57.64%	NA
Best in Class ESG	8.99%	16.12%	-0.31	4.87	0.54	-50.92%	NA
Commodity Pairs	26.81%	22.53%	0.90	10.61	1.18	-19.39%	0.0033%
Currency Pairs	1.67%	4.19%	0.44	9.73	0.083	-12.00%	0.0041%

Note: All statistics are reported as annualised figures except transaction costs which are the average % of notional trade value

5.1.1 Gradient Boosting Tree Strategy

The In-sample period for the GBT model, incorporating SHAP values analysis spanned from the beginning of 2014 to the end of 2017 with annual rebalancing of the portfolio. The model employed rolling windows, meaning each training period incorporated all available previous data, leveraging cumulative historical information. During the in-sample period, the portfolio demonstrated solid performance, achieving an annualised return of 17.66% and an annualised volatility of 18.60%, resulting in a Sharpe Ratio of 0.94. The relatively low Max Drawdown indicates that the portfolio experiences minimal fluctuations. Additionally, the portfolio's Kurtosis of 5.45, which exceeds the standard kurtosis of 3 for a normal distribution, suggests a distribution with more pronounced tails, implying a higher likelihood of extreme returns. The negative skewness of -0.11, while slight, indicates a marginal asymmetry in the distribution, with a subtle inclination towards negative values.

5.1.2 ESG Momentum Strategy

The in-sample period for the ESG momentum strategy ranges from 01/01/2008 until 31/12/2017, with the entirety of 2007 utilized to generate the ESG momentum signal. The sample period is thus one year shorter, due to the 12-month lag needed to construct the signal. Referring to the results presented in Table 1, the positive ESG momentum portfolio generated an annualised return of 5.79%, paired with a volatility of 20.45%, reflecting a strategy with balanced growth prospects tempered by substantial risk exposure. The strategy's Sharpe Ratio stood relatively low at 0.23, indicating positive risk-adjusted returns. Notably, the maximum drawdown was -57.64%, reflecting substantial but not extreme declines, although this decline is a reaction to the global financial crisis where equity markets around the world went down. The skewness of 0.09 indicates that returns are relatively symmetrically distributed, with a slight inclination towards positive outcomes. The higher kurtosis at 10 implies a likelihood of occasional outsized returns, which could mean infrequent but significant gains or losses. The strategy portrays moderate returns with the potential for both stable gains and significant fluctuations.

5.1.3 Best-in-Class ESG Strategy

From 2007 to 2017, during the in-sample period, the Best-in-Class strategy, to invest in the top 10% of ESG Score in each sector, achieved an annualised return of 8.34% and a remarkable annualised volatility of 17.85%. During this period, a maximum drawdown of -50.8% was recorded and the strategy generated a Sharpe Ratio of 0.47, indicating that the return is obtained with significant risk. It is important to note that these performance figures are calculated excluding transaction costs.

5.1.4 Dynamic Commodity Pairs Strategy

The in-sample period for the commodity pairs strategy spans from the beginning of 2007 through 2017. The strategy incorporates a 1-year rolling window, initiating live signals in early 2008. From 2014 until the end of the sample period, the strategy selects the top-performing 50% of pairs based on their risk-adjusted return contribution to the strategy. These pairs will then be the only pairs constituting the trading pool from 2014 until 2018. After this selection, the strategy delivered a respectable annual return of 17.8% with an annual volatility of 18.04%. The resulting Sharpe Ratio for the in-sample period stands at 0.99. Despite a drawdown of -13.95%, the strategy demonstrates resilience within its market context. Notably, the strategy exhibits a substantial kurtosis of 10.95, suggesting quite a high degree of tail risk in the return distribution. Additionally, the skew of 0.84 indicates that the strategy has a longer right tail and the potential for positive returns is higher than the potential for negative returns.

5.1.5 Commodity Terms of Trade FX Strategy

The in-sample testing period for this strategy spanned the beginning of 2007 until the end of 2017, the strategy requires 4 weeks (29 days) to generate trading signals and as a result, has zero returns until the first trading day at least 29 days from the beginning of the year. As can be seen from ***Table 1*** above the CToT strategy provided an annualised return of 1.67% as well as a positive risk-adjusted return, exhibiting a Sharpe Ratio of 0.02. During this period the strategy also experienced relatively low annualised volatility and moderate smaller maximum drawdown and positive skewness indicating a higher chance of extreme positive than negative returns.

After the in-sample testing period, the performance of each currency pair under the weighting scheme was evaluated in risk-adjusted terms, with currency pairs which contributed negatively by this metric (Sharpe Ratio) being excluded from out-of-sample trading, allowing greater weight to be allocated to the currency pairs which contributed positively. 4 currency pairs remained in the portfolio, the Brazilian Real, Israeli New Shekel, Nigerian Naira and Russian Ruble (all paired with the US dollar).

5.1.6 Combined Strategies

5.1.6.1 Hypothesis

As mentioned in the Introduction, it is assumed that the asset-weighted strategy will be the best performing. The strategy allocates the capital equally among the asset classes, leveraging the benefits of diversification. By assigning equal risk, to equities, commodities, and FX, rather than to the strategies, equities are not overweighted as in in the equal-weighted strategy. In this way, the strategy aims to mitigate the idiosyncratic risk of any single asset class dominating the portfolio's performance. This equal distribution of risk across varied asset classes is intended to harness the diversification effect, which can potentially lead to more stable and improved risk-adjusted returns over time. It does not rely on complex rebalancing algorithms or predictions about future volatility, which can be error-prone. By equally dividing risk among varied asset classes, it naturally adjusts to market changes, potentially offering a more robust performance in differing market conditions without the need for frequent adjustment. Furthermore, unlike the equal-weighted strategy, which disproportionately allocates 60% towards equities given its three individual equity strategies compared to one each for FX and commodities, the asset-weighted strategy offers a more balanced risk profile.

5.1.6.2 In-sample results

In shaping an effective portfolio allocation strategy, consideration of the individual strategies and their interactions within the broader market is crucial. The correlation among them underscores the potential benefits of prudent allocation, leveraging the diversification effects inherent in assets with low correlations. Notably, the presence of three individual strategies focused on ESG criteria in the Nordic market introduces a nuanced balance—offering diversification benefits while also necessitating a cautious approach to avoid overexposure to specific stocks. The inclusion of two alternative market strategies, FX (foreign exchange) and commodities, in our portfolio offers a diversification benefit. This is evident from their very low correlation of -0.0212 with each other, and the highest correlation with equities being only 0.0526. Such low correlation figures indicate that these assets move relatively independently of each other, presenting an opportunity to spread portfolio risk across different market segments. As expected, the correlation between the different Equity strategies, drawing returns from the same market, is quite high.

Table 2 - *Correlation Matrix - In sample period: 2014-2017*

	FX returns	ESG BIC	COM return	ESG MOM	ESG ML
FX returns	1.0000	0.0527	-0.0212	0.0478	-0.0052
ESG BIC	0.0526	1.000	0.0059	0.8925	0.7485
COM return	-0.0212	0.0059	1.0000	0.0126	-0.0171
ESG MOM	0.0478	0.8925	0.0126	1.0000	0.8369
ESG ML	-0.0052	0.7485	-0.0171	0.8369	1.0000

The subsequent section quantifies these considerations by examining both active and passive allocation approaches. It aims to offer a nuanced comprehension of the optimal weighting scheme for the diversified portfolio. Refer to Table 3 for the relevant statistics corresponding to each of the allocation strategies.

Table 3: *Combined Strategy in Sample Performance Statistics*

	Return	Volatility	Sharpe Ratio	Kurtosis	Max Drawdown	Skewness
Equal-Weighted	13.97%	10.48%	1.31	6.41	-11.19%	-0.23
Asset-Weighted	14.25%	9.16%	1.53	7.31	-7.57%	0.47
Risk-Parity	8.79%	6.57%	1.30	7.58	-6.14%	-0.19
Sharpe Optimizer	12.13%	7.24%	1.64	5.85	-5.68%	0.33
Benchmark	5.86%	5.03%	1.11	4.53	-6.84%	-0.19

Note: All statistics are reported as annualised figures

5.1.6.3 Equal-Weighted

In assessing the equal-weighted allocation strategy, it's clear that the various sub-strategies exhibit diversification even without optimization. While the strategy boasts a respectable Sharpe Ratio of 1.31 and annual returns of 13.97%, its performance is relatively mediocre in most metrics compared to the alternative allocations. As far as the distribution characteristics of the returns are concerned, a kurtosis of 6.41 and a skewness at -0.23 were determined. The maximum drawdown was calculated at -11.19 %.

5.1.6.4 Asset-Weighted

The asset-weighted approach distributes a larger allocation in the alternative markets, constituting 66% of the portfolio. Considering their negative correlations and different risk-return characteristics, the allocation performs significantly better than risk parity in most metrics. Limiting only 11% each to seemingly profitable equities strategies leads to a flatter and less volatile return profile when compared to the equity-weighted strategy. The Sharpe Ratio also has a higher value than the equal-weighted strategy.

5.1.6.5 Risk-Parity

As evidenced in Table 2, the risk-parity allocation strategy performs with mediocrity in terms of risk-adjusted returns when compared to the alternative allocations. It also has significantly lower annual returns. The predominant factor contributing to the resulting performance is the overweight position in the FX strategy, exceeding 53% of the total portfolio throughout the entire sample period. The remaining allocation was distributed relatively evenly among the other strategies for most of the period. This allocation was influenced by the FX strategy's low volatility during the in-sample period. However, it is crucial to recognize that this low volatility is accompanied by relatively subdued returns. Risk parity, relying on risk-based capital allocation, also encounters challenges in responding effectively to shifts in market sensitivity.

Despite being less favourable in terms of annual return, the risk-parity strategy still demonstrates a noteworthy Sharpe Ratio of 1.30. This Sharpe Ratio surpasses that of any individual strategy by quite some margin, underscoring the relative effectiveness of the risk-parity approach and the diversification effects of combining the different strategies.

5.1.6.6 Maximized Sharpe Ratio

Utilizing the Sharpe Optimizer approach, which seeks to enhance the Sharpe Ratio following principles similar to those advocated by Harry Markowitz, the strategy exhibits commendable performance. The strategy expectedly achieved the highest Sharpe Ratio of 1.64 after the optimization, indicative of considerable risk-adjusted returns, while also exhibiting lower volatility of 7.24% compared to the Equal-Weighted and Asset-Weighted strategies. This suggests a more stable performance with fewer fluctuations. The maximum drawdown is also the lowest at -5.68%, implying less probability of large downside fluctuations in portfolio value. Moreover, the positive skewness of 0.33 hints at a bias towards more frequent positive returns as opposed to negative ones. These statistics, collectively, underscore the effectiveness that could be obtained if one leverages the allocations to their strengths.

Interestingly, the weighting scheme adopted based on the in-sample period fully excludes 2 strategies – ESG BIC and ESG MOM, resulting in investments only being made into 3 of the strategies.

Table 6: *Combined Strategy Out of Sample Performance Statistics*

	Return	Volatility	Sharpe Ratio	Kurtosis	Max Drawdown	Skewness
Equal-Weighted	14.27%	12.82%	1.02	7.22	-21.26%	-0.59
Asset-Weighted	15.61%	10.63%	1.36	5.07	-11.29%	-0.15
Risk-Parity	15.07%	9.83%	1.42	5.29	-10.57%	-0.17
Sharpe Optimizer	17.38%	10.08%	1.61	4.95	-13.05%	-0.07
Benchmark	3.93%	7.45%	0.37	9.61	-16.65%	-0.22

Note: All statistics are reported as annualised figures

Table 4 - *Maximized Sharpe Ratio - weights*

FX	ESG BIC	COM	ESG MOM	ESG ML
52%	0%	24%	0%	24%

5.2 Out-of-Sample

The out-of-sample results present a comprehensive narrative about the efficacy of the in-sample tested portfolio optimization strategies. The correlation in the out-of-sample period between the individual strategies is different in certain aspects compared to the in-sample period. The ESG MOM portfolio showcases quite a different correlation to the COM and FX portfolios, with absolute differences of 0.0619 and 0.1221, demonstrating differences in the returns. The Same effect is noticeable for ESG BIC, with an absolute difference of 0.0559 and 0.1053 respectively. The ESG ML Showcases small differences in correlation, but the effect is not nearly as prevalent. Interestingly enough, The ESG strategies have a negative correlation towards the COM and FX portfolios, indicating even further diversifications than in the In-sample period, where only the ESG ML had a negative correlation.

Table 5 *- Correlation Matrix - Out-of-sample: 2018-2022*

	FX returns	ESG BIC	COM return	ESG MOM	ESG ML
FX returns	1.000	-0.0526	-0.0032	-0.0744	-0.0601
ESG BIC	-0.0526	1.0000	-0.0499	0.8988	0.6920
COM return	-0.0032	-0.0499	1.0000	-0.0493	-0.0381
ESG MOM	-0.0744	0.8988	-0.0493	1.0000	0.7897
ESG ML	-0.0601	0.6920	-0.0381	0.7897	1.0000

Table 5 provides a comprehensive overview of the performance metrics of the different investment strategies over the specified out-of-sample period. The equal-weighted strategy in the out-of-sample period had an annualised return of 14.27%, slightly higher than its in-sample return of 13.97%. However, the volatility also increased from 10.48% to 12.82%, and the Sharpe ratio fell from 1.31 to 1.02. The maximum drawdown of -21.26% was significantly higher than in the sample (-11.19%) and the skewness changed more negatively to -0.59.

This asset-weighted strategy performed better out-of-sample with an annualised return of 15.61%, compared to 14.25% in-sample. Volatility increased from 9.16% to 10.63%, and the Sharpe ratio worsened from 1.53 to 1.36. Maximum drawdown was less pronounced out-of-sample (-11.29% vs. -7.57%), and skewness decreased slightly to -0.15. The improvement in both returns and Sharpe ratio in the out-of-sample period suggests that this strategy is more adaptable to changing market conditions and provides a better balance between return and risk.

The annualised return for the risk-parity strategy was 15.07% out-of-sample, significantly higher than the 8.79% for the in-sample period. Volatility was higher at 9.83% compared to 6.57%, and the Sharpe ratio increased from 1.30 to 1.42. The maximum drawdown improved from -6.14% to -10.57%, with skewness remaining negative at -0.17. An increase in the return can also be observed here.

Out-of-sample, the Sharpe Optimizer achieved the highest annualised return of 17.38%, compared to 12.13% in the sample. Volatility was slightly higher at 10.08% compared to 7.24%, and the Sharpe ratio decreased from 1.64 to 1.61. The maximum drawdown was less pronounced at -13.05% compared to -5.68%, and skewness was slightly negative at -0.07.

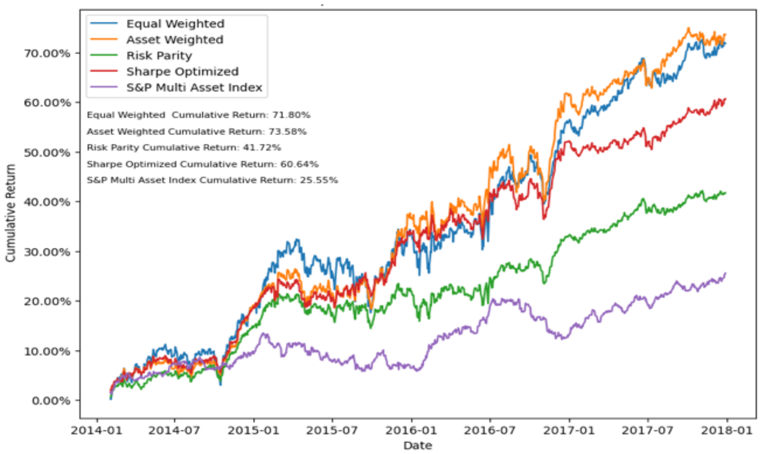
The equal-weighted strategy recorded marginal gains in returns at the expense of increased risk, while the asset-weighted strategy improved both returns and risk management. The risk parity strategy recorded a significant increase in returns and a balanced improvement in risk ratios. The Sharpe Optimiser strategy in particular saw a significant increase in returns and efficient risk-adjusted performance.

5.3 Discussion of Performance Results

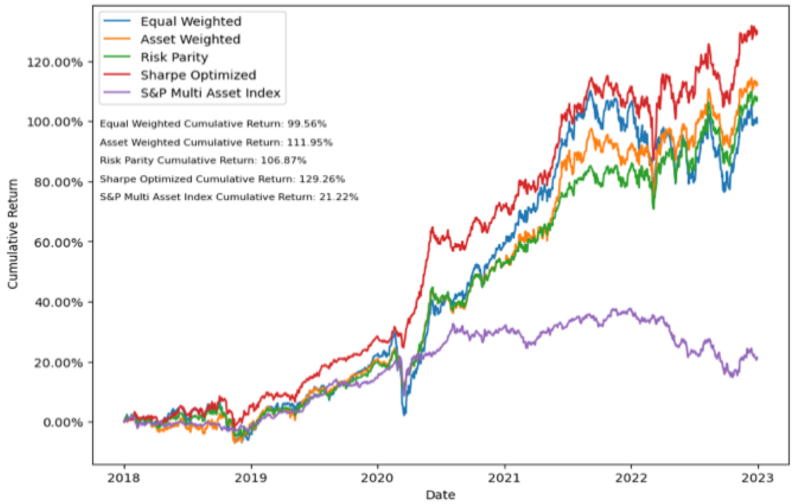
The varied outcomes from the portfolio strategies offer several notable points for discussion. The hypothesis that the asset-weighted portfolio will be the most optimal way is rooted in the equal distribution of risk across diverse asset classes like equities, commodities, and FX, and its diversification benefits. This diversification is firstly underscored by the correlation matrix showing the high correlation between the three equity strategies and the extremely low correlation between equities, FX and commodities around 0.05. While almost all portfolios increased in performance during the out-of-sample exciding the benchmark by far, the Sharpe optimizer stands out as the best performer due to its superior risk-adjusted returns and resilience as evidenced by its Sharpe Ratio and maximum drawdown metrics during in-sample and out-of-sample. It demonstrates that maximizing the Sharpe Ratio can lead to a strategy that not only performs well in-sample but also out-of-sample, a testament to its robustness. Nevertheless, in-sample and out-of-sample the second best-performing strategy was the asset-weighted portfolio, which also showed strong improvements, highlighting the effectiveness of diversification and balanced risk management. It stood out due to its superior Sharpe Ratio compared to the Equal-Weighted and Risk-Parity strategies, while also limiting volatility through a balanced allocation across various asset classes. It showed notable improvements from the in-sample period, achieving an annualized return of 15.61% and an improved Sharpe ratio of 1.36. This strategy demonstrated adaptability and an enhanced balance between return and risk, suggesting it was better suited to the changing market conditions of the out-of-sample period.

Some notable events occurred during the investment periods, where the market conditions varied greatly from the in-sample period to the out-sample period. As indicated by the graph below, the in-sample period was relatively stable with not a lot of short crashes and a relatively stable increase over time. However, the out-sample period graph tells a different case, where extreme spikes happen during notable times. After 2020 there was a large dip in performance, mostly explained by the volatile period of COVID-19. However, The Sharpe Optimized portfolio does not see a similar spike as the others. This can be explained by the Weighting scheme of the Sharpe Optimized portfolio, where 52% of the funds are allocated to the FX strategy. This indicates that the FX-strategy performs well during times of market uncertainty, and further reinforces the diversification benefits. The same phenomenon is however not present in 2022, another year of high market uncertainties, where the first increases in risk-free rate started. The Graph also highlights the volatile nature of the year 2022, where the returns fluctuate a lot, presumably because of the war in Ukraine, which had dramatic effects on the market. The chosen Benchmark does offer a much safer investment strategy than anticipated and seems to consistently offer safer but lower returns.

In-sample Portfolio performance



Out-of-sample portfolio performance



Comparing the change from in-sample and out-of-sample performance, the Risk-Parity strategy showed the highest improvement. It had a marked increase in its annualized return, going from 8.79% in-sample to 15.07% out-of-sample. Additionally, its Sharpe Ratio improved from 1.30 to 1.42, reflecting better risk-adjusted performance despite increased volatility and a higher maximum drawdown in the out-of-sample period.

In conclusion, the asset-weighted strategy, with its balanced risk distribution and more realistic adaptability, seems to offer the most compelling case for a robust portfolio management approach. Besides, the Sharpe Optimizer strategy also stands out for its efficient risk-adjusted returns and stability but is trickier to integrate into a retail portfolio.

5.4 Regression

This section presents a regression analysis performed on the investment strategies against a specified benchmark, Fama French 5 Factors + Momentum (FF5FM). As our portfolio is quite diversified and covers different areas of the market, the chosen Benchmark was the S&P U.S. Balanced Multi-Asset Index. This benchmark is reflective of the asset classes targeted by the strategies: equities, FX, and commodities, but also includes fixed income. The benchmark was chosen as it is a fair representation of the returns an investor could generate passively by investing in a multi-asset index.

For this analysis, monthly returns data for each strategy are adjusted by the risk-free rate to compute excess returns, which are then subjected to a regression against the excess returns of the benchmark and FF5FM. This methodology facilitates a thorough investigation of the alpha and beta coefficients of each strategy. The alpha coefficient offers insight into the performance of the strategy beyond the market's influence, while the beta coefficient quantifies the sensitivity of the strategy's returns to market fluctuations, serving as an indicator of systematic risk.

Table 7: Regression Results **- *Fama French Factors + Momentum***

	In-Sample: 2014-2017			
	Equal weight	Asset Weighted	Risk parity	Max Sharpe
Alpha (a)	14.65%**	16.15%**	8.85%**	12.56%**
	(0.000)	(0.000)	(0.000)	(0.000)
Market (b)	0.26	0.14	0.16	0.11
	(0.270)	(0.585)	(0.302)	(0.550)
Size	0.28**	0.33**	0.22**	0.27**
	(0.000)	(0.000)	(0.000)	(0.000)
Value	-0.35**	-0.34**	-0.19**	-0.17**
	(0.000)	(0.000)	(0.000)	(0.000)
Profitability	-0.0844	-0.09	0.02	0.09
	(0.160)	(0.121)	(0.623)	(0.053)
Investment	0.13	0.62**	0.046	0.41**
	(0.069)	(0.000)	(0.312)	(0.000)
Momentum	-0.0587	0.13**	-0.01	0.16**
	(0.028)	(0.000)	(0.780)	(0.000)
R^2	0.138	0.179	0.131	0.170
Information Ratio	1.81	1.89	1.69	1.93
Tracking Error	0.081	0.085	0.052	0.065

Notes: p-values in (); **,* denote statistical significance at 1 and 5% level, respectively. Alpha ,Information ratio and Tracking error are given in annual terms.

Within the in-sample period, the Asset Weighted portfolio has the highest alpha of 16.15%, with an Information ratio of 1.89 and a Tracking error of 0.085, showcasing a strong performance on key metrics, while also having the highest R-squared, 0.180, indicating the most correlation with FF5F+Momentum. The Sharpe Optimized Portfolio, not surprisingly, performs the highest in terms of Information Ratio at 1.93 while having an alpha of 12.56%. all the portfolios have a p-value below 0.05 for the Alpha, indicating that the results are statistically significant. In terms of the other factors, Size & Value have all a p-value below 0.05 across the board and are hence statistically significant. The Size factors are positive for the portfolios, which indicates that the portfolios reflect a tilt or preference towards small-cap companies or stocks. Conversely, the negative relationship with the value factor indicates that the portfolios favour growth over value stocks. The asset-weighted portfolio and Sharpe Ratio-optimized portfolio have significantly positive values for the Momentum factor, indicating that during the sample period, those strategies capitalized on a continuation of existing market trends. The same relationship could be seen with the Investment factor, implying a preference for or better performance with firms that are actively investing (high-investment firms). In terms of the Market factor, none of the P-values are statistically significant and hence no conclusions can be drawn.

Table 8 : Regression Results **- *Fama French Factors + Momentum***

	Out-of-Sample: 2018-2022			
	Equal weight	Asset Weighted	Risk parity	Max Sharpe
Alpha (a)	10.74%*	12.97%**	12.05%**	14.19%**
	(0.026)	(0.008)	(0.005)	(0.002)
Market (b)	0.99**	0.43**	0.50**	0.45**
	(0.000)	(0.036)	(0.006)	(0.019)
Size	0.43	0.25	0.2341	0.16
	(0.012)	(0.133)	(0.115)	(0.305)
Value	0.06	-0.05	-0.0041	-0.15
	(0.662)	(0.750)	(0.974)	(0.264)
Profitability	0.25	0.07	0.1819	0.04
	(0.205)	(0.711)	(0.291)	(0.808)
Investment	-0.17	0.04	-0.0512	0.07
	(0.405)	(0.849)	(0.771)	(0.708)
Momentum	-0.07	-0.04	-0.0522	-0.02
	(0.563)	(0.733)	(0.611)	(0.894)
R^2	0.499	0.161	0.265	0.188
Information Ratio	1.16	1.39	1.47	1.62
Tracking Error	0.093	0.093	0.082	0.088

Notes: p-values in (); **,* denote statistical significance at 1 and 5% level, respectively. Alpha, Information Ratio and Tracking error are given in annual terms.

During the out-of-sample period, the regression analysis exhibits similarities in Factor values but demonstrates notable differences, particularly in terms of p-values. The Sharpe-Optimized portfolio had the highest alpha of 14.19%, followed by the Asset Weighted portfolio of 12.97%. Interestingly, the Equal Weighted portfolio shows a much higher correlation with an R-square of 0.499 and the same trend is also noticeable for the rest of the portfolios. The alpha is, similarly to the in-sample period, statistically significant. An interesting change is that all portfolios now showcase a p-value below 0.05 for the market factor, showcasing a statistically significant relationship. This means that, as the Market factor is positive across the board, the portfolios tend to move in the same direction as the Benchmark. For the equal-weighted portfolio, The Factor is 0.9928, meaning that the returns of the strategy are very closely aligned with the Benchmark. Most of the other factors, Size, Value, Profitability and Momentum, do not have statistically significant P-values for the out-of-sample period, barring the Size Factor for the Equal Portfolio. Hence, no real conclusions from these factors can be drawn for the Out-of-sample period, regardless of their factor value.

In summary, all of the portfolios showcase a high positive alpha towards the Benchmark, both during the In-sample-period and Out-of-sample period. The high Information Ratio also tells us that our portfolios consistently outperform the given Benchmark, While the relatively High Tracking error tells us that the returns of the portfolio's returns are not closely aligned with the Benchmark. As described in the previous section, that result is not surprising, as the chosen Benchmark has, over both the in-sample and out-of-sample period, a much lower Annualised Volatility and Annualised Returns. As the FF5 + Momentum factors p-value changes in between the periods, no real relationship can be consistently drawn across the entire investment period, and further analysis combining or analysing different periods needs to be conducted to further validate the claims.

6 Limitations

A key challenge investors encounter stems from the costs of operating an active strategy. Transaction costs vary based on factors such as monthly trade volumes and whether the investor is categorised as retail or professional, with the latter often benefiting from lower fees. The individual strategies GBT, CToT and Commodity Pairs Trading consider different transaction costs based on their asset class and the nature of their trades aside from ESG

Momentum and ESG Best-in-Class, which may lead to an overestimation of actual returns for these two strategies. Further, the strategy is for simplicity's sake rebalancing each weighting scheme to the weight of that year. This means that winners get punished and could skew the results negatively. It is however also possible that strategies that fluctuate highly in a certain year do not get punished for it, as the fluctuation is not fully encapsulated by the weighting scheme.overall, the effect should presumably skew returns upwards, as Transaction costs for these trades are not included. Besides explicit costs, investors must account for implicit costs such as the bid-ask spread and slippage/latency, typically more significant for professionals due to the larger size of their trades. We use the end-of-day/week/month closing price in our analysis. Still, it is important to note that this price might not always be achievable in real trading scenarios, where buying and selling occur at ask and bid prices, respectively. High market volatility, like that observed during the beginning of the COVID-19 pandemic, can widen spreads, especially for illiquid assets, and even minimal spreads can accumulate significantly over time, impacting performance.

Additionally, liquidity can pose a major challenge in the commodity futures strategy, particularly with distant contracts that are most liquid during monthly rollovers. Investors needing immediate liquidity to act on trading signals might find this strategy underestimates the impact of illiquidity. The commodity futures market is also evolving rapidly, with its financialization detailed in a separate report. As this market expands, it is unclear how the risk premium will be affected. Similarly, in the commodity terms of trade strategy, the tendency to trade more in emerging, commodity-dependent currencies, which can be less liquid and have higher bid-ask spreads, can significantly increase transaction costs as well as have the potential to limit the scalability of the strategy in the case of, for example, a large institution trading high volumes in currencies with lower market liquidity.

Furthermore, the commodity strategy mandates the maintenance of a margin account to facilitate the use of leverage. While the strategy allows for leveraging futures contracts, it demands a substantial capital commitment due to the sizable notional value associated with certain contracts. For example, strategies employing a long-short approach require a margin account for stock borrowing, making them challenging for retail investors. This difficulty arises not only from the risks and financial savvy needed for managing gross exposures exceeding 100% but also due to the same substantial capital requirements. Portfolio rebalancing, essential for achieving optimal asset allocation, involves buying and selling stocks and futures. While stock prices typically range from $10 to $1,000, dealing with commodity futures, where a single contract can be worth up to $100,000, complicates rebalancing. Therefore, achieving precise asset allocation is mostly feasible for institutional-sized portfolios.

Moreover, the adage “past performance is not indicative of future returns” is particularly relevant in finance and more applicable than ever. The analysis period began post-global financial crisis and post-Eurozone debt crisis, encompassing a notably strong decade for equities and the post-pandemic surge linked with the ECB’s unconventional monetary policies, including quantitative easing and negative interest rates.

Expanding on this point, portfolios like Risk-parity and Maximzed Sharpe are constructed using historical returns, leading to a bias that favors portfolios with exceptional past performance. Consequently, this bias results in certain weightings that exclude entire portfolios, undermining the diversification advantages. For instance, in the Maximum Sharpe portfolio, no investments are allocated to ESG BIC and ESG MOM, resulting in only 3 strategies being invested in.

In addition, an overfitting bias in the in-sample period is present for the Maximized Sharpe Ratio strategy as the weights of the strategies were fitted to the historical data on which they were tested, resulting in a model that is intricated and specific to the idiosyncrasies of this particular data set. However, this bias is not present in the out-of-sample period, as the weighting scheme is the same as the one optimized for the in-sample period.

The forward-looking bias in our ESG scores is another notable limitation. This bias emerges from the reliance on contemporary ESG data that may not have been accessible or fully developed during the historical periods under examination, which can lead to artificially deflated or inflated performance results, painting an overly optimistic picture of the investment strategies based on this data. As a result, the backtesting may inadvertently incorporate insights and assessments based on information that, at the time, was not available to investors or the public.

Survivorship bias is an important factor to consider in our composite strategy, particularly for the equity strategies that utilize the OMX Nordic Large Cap index. Such Indices undergo periodic revisions to their constituent stocks, typically adding (subtracting) stocks due to high (low) performance or due to higher (lower) market capitalization. This process can induce survivorship bias, presenting an inflated historical performance by only reflecting currently successful stocks and omitting those eliminated due to poor performance. In financial analysis, failure to acknowledge this bias can result in a distorted perception of performance, leading to potentially flawed investment decision-making predicated on partial or biased data.

The Equity strategies also pose a certain risk in terms of currency fluctuations, as the chosen Market incorporates 4 different types of Currencies – The Swedish Crown, The Norwegian Crown, The Danish Crown and The Euro. Typically, one would assume a cost of Hedging these types of risks, but due to the complication of implementing it, it has been excluded, which means that the returns, in a realistic scenario, would further be decreased.

7 Results/ Conclusions

This work was a comprehensive attempt to evaluate the performance and profitability of combining different investment strategies within a multi-asset portfolio, particularly in the Nordic market. Incorporating commodities, foreign exchange, and the three novel ESG approaches, Gradient Boosting Tree, ESG Momentum, and Best-in-Class, was noteworthy. Each of these strategies brought distinct features and performance trends to the overall portfolio, highlighting the intricacies of multi-asset investment. The analysis uncovered that although conventional asset categories like equities and commodities are fundamental to portfolio building, integrating ESG factors as non-financial components could enhance risk-adjusted returns.

For each strategy, the empirical analysis yielded different results. The asset-weighted approach, which heavily invested in alternative markets, outperformed the risk parity method across most metrics. This indicates the advantages of diversifying across asset classes. The Maximum Sharpe Ratio strategy demonstrated its effectiveness by delivering the highest risk-adjusted returns. This confirms the notion that a well-balanced multi-asset strategy can generate above-average portfolio performance. However, strategies such as risk parity provided stable but lower total returns. This highlights the trade-off between managing risk and maximizing returns.

The research confirmed the original hypothesis: diversification across commodities and currencies and the inclusion of ESG strategies can improve a portfolio's risk-adjusted returns. This finding is central to understanding the construction of multi-asset portfolios in today's markets, particularly given the increasing importance of ESG factors in investment decisions.

To advance the study of multi-asset portfolio management, a wider and more diverse range of research is necessary. Examining various global markets, not limited to the Nordic region studied here, could yield valuable insights into the adaptability and variability of these strategies in different economic contexts. Additionally, exploring the inclusion of emerging market dynamics into these portfolios presents an interesting opportunity, given the potential for growth and unique risk-return profiles of these markets. Such research could investigate how emerging market assets interact with traditional assets like bonds, equities, and commodities, potentially revealing novel strategies for risk management and diversification. Moreover, incorporating bonds into multi-asset frameworks would provide a more comprehensive view of portfolio construction and a deeper understanding of how different asset classes can synergize to enhance portfolio resilience and performance in varying market conditions.

To summarise, this paper demonstrates the significant benefits of a multi-asset approach to portfolio management, with a focus on the integration of ESG factors. The findings provide investors with clear guidance on the benefits of diversified, risk-adjusted strategies. These findings highlight the effectiveness of multi-asset portfolios in achieving balanced performance in different market conditions and emphasize their relevance in the broader context of global investment strategies.

8 References

Albuquerque, Rui, Yrjö Koskinen, and Chendi Zhang. 2019. "Corporate Social Responsibility and Firm Risk: Theory and Empirical Evidence." *Management Science* 65 (10): 4451–69. doi:10.1287/mnsc.2018.3043.

Arezki, M.R., Pattillo, M.C.A., Quintyn, M.M.G., and Zhu, M. 2012. "Commodity Price Volatility and Inclusive Growth in Low-Income Countries." International Monetary Fund.

Bennani, Leila, Théo Le Guenedal, Frédéric Lepetit, Lai Ly, Vincent Mortier, Thierry Roncalli, and Takaya Sekine. 2018. "How ESG Investing Has Impacted the Asset Pricing in the Equity Market." *Social Science Research Network*, January. doi:10.2139/ssrn.3316862.

Berg, Florian, Julian F Kölbel, and Roberto Rigobón. 2022. "Aggregate Confusion: The Divergence of ESG Ratings." *Review of Finance* 26 (6): 1315–44. doi:10.1093/rof/rfac033.

Bilson, John F. O., and David A. Hsieh. 1987. "The Profitability of Currency Speculation." *International Journal of Forecasting*3 (1): 115–30. doi:10.1016/0169-2070(87)90082-3.

Bollerslev, Tim, Benjamin Hood, John Huss, and Lasse Heje Pedersen. 2018. "Risk Everywhere: Modeling and Managing Volatility." *The Review of Financial Studies* 31 (7): 2729–73. doi:10.1093/rfs/hhy041.

Brzenk, Phillip, Liu, Berlinda and Tianyin Chen. 2020. "Indexing Risk Parity Strategies." S&P Dow Jones Indices, October 14.

Cashin, Paul, Luis Felipe Céspedes, and Ratna Sahay. 2004. "Commodity Currencies and the Real Exchange Rate." *Journal of Development Economics* 75 (1): 239–68. doi:10.1016/j.jdeveco.2003.08.005.

Chen, Tianqi, and Guestrin, Carlos. 2016. "*XGBoost: A Scalable Tree Boosting System*." In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, 785–794. doi:10.1145/2939672.2939785

Chen, Zhiwu, and Peter J. Knez. 1995. “Measurement of Market Integration and Arbitrage”. Review of Financial Studies, 8: 287-325.

Cheung, Yin-Wong, Menzie David Chinn, and Antonio García Pascual. 2005. “Empirical Exchange Rate Models of the Nineties: Are Any Fit to Survive?” *Journal of International Money and Finance* 24 (7): 1150–75. doi:10.1016/j.jimonfin.2005.08.002.

Collier, Paul, and Benedikt Goderis. 2012. “Commodity Prices and Growth: An Empirical Investigation.” *European Economic Review* 56 (6): 1241–60. doi:10.1016/j.euroecorev.2012.04.002.

Dimson, Elroy, Oguzhan Karakas, and Xi Li. 2015. "Active Ownership." Review of Financial Studies 28 (12): 3225-3268.

Eccles, Neil Stuart, and Suzette Viviers. 2011. “The Origins and Meanings of Names Describing Investment Practices That Integrate a Consideration of ESG Issues in the Academic Literature.” *Journal of Business Ethics* 104 (3): 389–402. doi:10.1007/s10551-011-0917-7.

Eccles, Robert G., Ioannis Ioannou, and George Serafeim. 2014. “The Impact of Corporate Sustainability on Organizational Processes and Performance.” *Management Science* 60 (11): 2835–57. doi:10.1287/mnsc.2014.1984.

Erb, Claude B. and Harvey, Campbell R. 2006. “The Tactical and Strategic Value of Commodity Futures”. SSRN Electronic Journal.

Erten, Bilge, and José Antonio Ocampo. 2013. “Super Cycles of Commodity Prices since the Mid-Nineteenth Century.” World Development, 44 (April): 14–30. doi:10.1016/j.worlddev.2012.11.013.

European Securities and Markets Authority (ESMA). 2021. "Report on Trends, Risks, and Vulnerabilities." European Securities and Markets Authority.

Eurosif. 2018. "European SRI 2018 Study." https://www.eurosif.org/wp-content/uploads/2021/10/European-SRI-2018-Study.pdf.

Fama, Eugene F. 1984. “Forward and Spot Exchange Rates.” *Journal of Monetary Economics* 14 (3): 319–38. doi:10.1016/0304-3932(84)90046-1.

Filippou, Ilias, and Mark P. Taylor. 2017. “Common Macro Factors and Currency Premia.” *Journal of Financial and Quantitative Analysis* 52 (4): 1731–63. doi:10.1017/s0022109017000424.

Friede, Gunnar, Timo Busch, and Alexander Bassen. 2015. “ESG and Financial Performance: Aggregated Evidence from More than 2000 Empirical Studies.” *Journal of Sustainable Finance & Investment* 5 (4): 210–33. doi:10.1080/20430795.2015.1118917.

Froot, Kenneth A., and Richard H. Thaler. 1990. “Anomalies: Foreign Exchange.” *Journal of Economic Perspectives* 4 (3): 179–92. doi:10.1257/jep.4.3.179.

Galema, Rients, and Dirk Gerritsen. 2022. “ESG Rating Score Revisions and Stock Returns.” *Social Science Research Network*, January. doi:10.2139/ssrn.4218969.

Galema, Rients, Auke Plantinga, and Bert Scholtens. 2008. “The Stocks at Stake: Return and Risk in Socially Responsible Investment.” *Journal of Banking and Finance* 32 (12): 2646–54. doi:10.1016/j.jbankfin.2008.06.002.

Gloßner, Simon. 2017. “ESG Risks and the Cross-Section of Stock Returns.” *Social Science Research Network*, January. doi:10.2139/ssrn.2980917.

Gorton, Gary, and K. Geert Rouwenhorst. 2006. “Facts and Fantasies about Commodity Futures.” Financial Analysts Journal 62 (2): 47–68. doi:10.2469/faj.v62.n2.4083.

Hallez, Emile. 2022. “Variability in ESG Ratings Makes a Murky Picture for Investors.” *ESG Clarity*. August 19. https://esgclarity.com/variability-in-esg-ratings-makes-a-murky-picture-for-investors/.

Hartley, Jonathan. 2020. “Interest Rate Momentum Everywhere across Global Yield Curves.” *Social Science Research Network*, January. doi:10.2139/ssrn.3542813.

Harvey, Campbell R., Edward Hoyle, Russell Korgaonkar, Sandy Rattray, Matthew Sargaison, and Otto Van Hemert. 2018. “The Impact of Volatility Targeting.” *Social Science Research Network*, January. doi:10.2139/ssrn.3175538.

Huck, Nicolas, and Komivi Afawubo. 2014. “Pairs Trading and Selection Methods: Is Cointegration Superior?” Applied Economics, 47 (6): 599–613. doi:10.1080/00036846.2014.975417.

IMF. 2023. "Groups and Aggregates." World Economic Outlook. https://www.imf.org/en/Publications/WEO/weo-database/2023/April/groups-and-aggregates.

Kaiser, Lars. 2020. “ESG Integration: Value, Growth and Momentum.” Journal of Asset Management 21 (1): 32–51. doi:10.1057/s41260-019-00148-y.

Kempf, Alexander, and Peer Osthoff. 2007. "The Effect of Socially Responsible Investing on Portfolio Performance." European Financial Management 13 (5): 908–22. https://doi.org/10.1111/j.1468-036x.2007.00402.x.

Khan, Mozaffar, George Serafeim, and Aaron Yoon. 2016. "Corporate Sustainability: First Evidence on Materiality." The Accounting Review 91 (6): 1697-1724.

Krauß, Christopher. 2016. "STATISTICAL ARBITRAGE PAIRS TRADING STRATEGIES: REVIEW AND OUTLOOK." Journal of Economic Surveys, 31 (2): 513–45. doi:10.1111/joes.12153.

Kroencke, Tim Alexander, Felix Schindler, and Andreas Schrimpf. 2011. "International Diversification Benefits with Foreign Exchange Investment Styles." *Social Science Research Network*, January. doi:10.2139/ssrn.1790764.

Lundberg, S. (2017, May 22). A Unified Approach to Interpreting Model Predictions. arXiv.org. https://arxiv.org/abs/1705.07874v2

Markowitz, Harry M. 1952. "Portfolio Selection." The Journal of Finance, 7 (1): 77. doi:10.2307/2975974.

Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen. 2011. "Time Series Momentum." *Social Science Research Network*, January. doi:10.2139/ssrn.2089463.

Nagy, Zoltán T., Douglas G. Cogan, and Dan Sinnreich. 2013. "Optimizing Environmental, Social and Governance Factors in Portfolio Construction: Analysis of Three ESG-Tilted Strategies." *Social Science Research Network*, January. doi:10.2139/ssrn.2221524.

Nagy, Zoltan, Kasam, Atlaf and Linda-Eling Lee. 2016. "Can ESG Add Alpha?" SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3175538.

Ng, Victor, and Stephen Craig Pirrong. 1994. "Fundamentals and Volatility: Storage, Spreads, and the Dynamics of Metals Prices." *The Journal of Business* 67 (2): 203. doi:10.1086/296630.

NYU Stern. Uncovering the Relationship by Aggregating Evidence from 1,000 Plus Studies Published between 2015 – 2020. New York University Stern School of Business 2021, URL:https://www.stern.nyu.edu/sites/default/files/assets/documents/ESG%20Paper%20Aug%202021.pdf

Ouaknine, Yannick, Nimit Agarwal, Aiswarya Sankar, Lorna Lyett, and Virgile Haddad. 2019. "ESG Rating and Momentum: Do Stocks with Improving ESG Ratings Represent Potential Upside for Investors?" Societe Generale Cross Asset Research, Equity Research, March 5.

TangKe, and XiongWei. 2012. “Index Investment and the Financialization of Commodities.” Financial Analysts Journal, 68 (6): 54–74. doi:10.2469/faj.v68.n6.5.

Vidyamurthy, Ganapathy. 2004. Pairs Trading: Quantitative Methods and Analysis. New Jersey: John Wiley & Sons, Inc.doi:10.1007/s40685-014-0015-7.

Vojtko, Radovan, and Matúš Padyšák. 2019. “Quant’s Look on ESG Investing Strategies.” *Social Science Research Network*, January. doi:10.2139/ssrn.3504767.

Von Wallis, Miriam, and Christian Klein. 2014. “Ethical Requirement and Financial Interest: A Literature Review on Socially Responsible Investing.” *Business Research* 8 (1): 61–98. Bossaerts, Peter, Richard C. Green. 1989. “A General Equilibrium Model of Changing Risk Premia: Theory and Tests”. The Review of Financial Studies, 2(4): 467–493.

Ouyang, Liwen. 2020. *"Seeking Signals from ESG Data*." Bloomberg Enterprise Quants. 2020, URL: https://data.bloomberglp.com/professional/sites/10/798791_SeekingSignalsFromESGData.pdf

A Work Project, presented as part of the requirements for the Award of a Master’s degree in Finance from the Nova School of Business and Economics.

Machine Learning in ESG Investing: Predictive analysis for stock performance

Joel Benjamin Lidén

Work project carried out under the supervision of:

Nicholas H. Hirschey

20-12-2023

Abstract

This strategy explores the use of ESG metrics within machine learning frameworks, particularly Gradient Boosting Trees and SHAP-value analysis, to predict stock performance in the Nordic Markets. Focusing on data from 2007-2022, it examines the efficiency of ESG metrics in forecasting Active Return. Utilizing the model, it reveals that long portfolios, especially those informed by SHAP analysis, consistently outperform the OMX Nordic Large cap Index, while short portfolios show underperformance. The study highlights the potential of machine learning in enhancing ESG-focused investment strategies, suggesting the need for broader datasets and diverse market analysis for more robust and comprehensive investment insights.

Keywords**:** *Finance, Financial Markets, Data Analysis, Equity ESG,Long Short, Sustainability, Quantitative Strategies, Nordic Stock Market, Sustainable Investment,Machine Learning, Gradient Boosting Tree,XGBoost, SHapley Additive exPlanations*

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. Introduction

In the last decade, Environmental, Social, and Governance (ESG) considerations have gained significant traction in investment decision-making. This trend reflects a growing recognition among conscientious investors of the importance of ESG metrics in evaluating a company's long-term financial health. Although ESG is a relatively new phenomenon, its components - particularly governance and social metrics - have long been vital in assessing corporate sustainability and resilience. The focus on the environmental aspect has intensified in recent years. Yet, it is crucial to understand that all three ESG metrics collectively offer a comprehensive view of a company's financial well-being.

The intersection of machine learning and investment strategies has its roots in the 1990s, where its incorporation into ESG investing has risen in the last decade yet is still in an infant phase. This paper embarks on an exploration of this untouched area, aiming to improve portfolio performance utilizing Gradient Boosting Trees (GBT). Furthermore, the most important features are selected using Shapley Additive exPlanations (SHAP), a game theoretic approach used to explain the output of a machine learning model.

This study aims to construct a viable trading strategy by utilizing ESG metrics to predict Active Return, using OMX Nordic Large Cap Index (OMXNLCI) as the benchmark, focusing on stocks within this index from 2007 to 2022. numerous ESG metrics are gathered annually, and the analysis involves a total holding period of 8 years with annual rebalancing.

2. Literature review

ESG, first introduced in the United Nations Principles for Responsible Investing in 2004, has gained substantial importance in recent years, sparking extensive debates on its application and relevancy. While it is generally agreed upon as critical for reaching net zero targets, there is ongoing debate about whether its application can also generate excess return, lower volatility, or both.

To clarify the mixed role of ESG, a seminal study by Friede, Busch, & Bassen (2015), which reviewed over 2000 different studies on ESG, established a compelling link between high ESG ratings and superior financial performance across global markets. Their findings highlight a positive correlation, indicating that companies excelling in ESG practices tend to demonstrate superior financial performance. This suggests that companies prioritizing environmental sustainability, social responsibility, and effective governance consistently outperform others.

This view is further reinforced by a comprehensive meta-analysis conducted by NYU Stern (2021), examining over 1,000 studies, uncovering the relationship between ESG and financial performance. The analysis found positive correlations between ESG performance and various aspects such as operational efficiencies, stock performance, and lower cost of capital. Notably, it revealed that the financial benefits of ESG become more apparent over longer time horizons, with studies focusing on the long-term being 76% more likely to find positive or neutral results.

Before the evolving studies on the impact of ESG, the field of machine learning had already been undergoing significant advancements. Key areas of development include time series forecasting, sentiment analysis, feature engineering, deep learning, and ensemble methods. The field saw notable developments with the introduction of Gradient Boosting Trees (GBT) by Jerome H. Friedman in 1999. Following this, the XGBoost library, developed by Tianqi Chen and Carlos Guestrin in 2016, emerged as an efficient and robust implementation of GBT, and is recognized for its effectiveness in complex predictive modeling tasks. Further enriching the interpretability of these sophisticated models, the incorporation of cooperative game theory, notably the Shapley value approach develpoed by Scott M. Lundberg and Su-In Lee (2017), has provided an insightful framework for understanding model predictions and their contributing factors.

Merging ESG metrics with machine learning, specifically in applying tree-based models to correlate ESG metrics with Active Return of stocks, is an underexplored area. Notably, there is a significant gap in comprehensive academic research in this specific area. The only notable reference is a limited analysis by Bloomberg in 2020, which showcased promising outcomes by outperforming the Russel 5000 and S&P 500. Apart from this, research illustrating the capability of machine learning methods to predict financial performance indicators such as ROE & ROA using ESG and other economic indicators, such as De Lucia, Pazienza, and Bartlett (2020) suggests a potential link between ESG variables and financial performance.

3. Strategic Reasoning

In the swiftly evolving domain of ESG, the significance of incorporating ESG factors into investment decisions is becoming increasingly recognized. However, this growth has introduced challenges, particularly in the realm of ESG scoring. With over 600 agencies utilizing a variety of methodologies, many of which lack transparency, investors find themselves navigating through a complex landscape of ESG assessments (Esma,2021). This ambiguity in ESG scoring is exemplified by cases such as Tesla's removal from the S&P 500 ESG Index while ExxonMobil, known for its problematic ESG practices, remains. Such decisions, influenced by diverse rating scales and methodologies, underscore the challenges in standardizing ESG evaluation (ESG Clarity 2022).

The lack of uniform regulations in the ESG field exacerbates these issues. The varied approaches of agencies, based on different and often unclear models, hinder investors from making reliable ESG assessments. This not only complicates investment decisions but also opens doors for more transparent, consistent ESG models. A novel approach, potentially through machine learning, could offer more clarity and dependability in ESG evaluations.

Machine learning, with its ability to process complex data, is particularly suited for identifying links between ESG metrics and financial performance. This is crucial in regions like the Nordic markets, pioneers in ESG innovation, where a rich dataset facilitates machine learning analysis (ESGinvesting 2020). Hence, applying machine learning to model ESG metrics on financial performance could be a key strategy to overcome current ESG investment challenges, enhancing transparency and accuracy in ESG ratings.

Focusing on the Nordic market, the goal is to predict Active returns based on ESG metrics to outperform the OMX Nordic Large Cap Index(OMXNLCI). By employing machine learning, this model will analyse ESG metrics in relation to stock performance, providing a more detailed and insightful analysis. This targeted approach in the Nordic region, known for its advanced ESG practices, could establish a new standard for ESG-informed investing, offering a deeper understanding of how ESG metrics affect financial outcomes.

4. Data and Methodology

4.1 Data

The dataset used for analysis has been retrieved solely from Refinitiv Platform, which provides financial data on the global market. The data ranges from 2007 to 2022, where the period 2007 to 2013 compiles the training period, and the years 2014 to 2017 is used as the in-sample period, and the period 2018 to 2022 compiles the out-sample period. Reasoning for the selected periods stems from the availability of data in combination with a split where there is sufficient training data to train the model on, while also ensuring an adequate in-sample and out-of-sample period for thorough analysis of the results.

Further, the period from 2013-2022 incorporates rolling windows, meaning that the model is trained on all currently available data. The chosen universe is the OMX Nordic large cap (OMXLC), and the dataset consists of ESG metrics for each individual stock as the X variables.

The justification for choosing the Nordic market stems from the fact that Nordic companies are regarded as leaders when it comes to sustainability and that the availability of ESG data is high from early on compared to other markets.

For the Y variable, each stock's annual Active Return from the OMX large cap Nordic index(OMXLCGI) is calculated and selected. The justification for using Active Return rather than purely using a stock's normal return is that the model is tasked with finding outliers compared to peers, rather than just stock returns relative performance. This goes in hand with the theoretical justification that companies in the high ESG spectrum should outperform peers and companies in a similar environment.

4.2 Data handling

Refinitiv provides annual ESG data for all companies in OMX Nordic, ranging back from 2006 up to the current date. Currently, there are 729 ESG metrics available on Refintiv for the OMXLC universe, with metrics updating annually. The initial dataset consisted of 729 ESG metrics and the final dataset consisted of 353 ESG metrics. Exclusion of ESG metrics stems mainly from 100% missing rate or irrelevancy, to make sure that the comparable analysis between stocks is accurate. Working with ESG data is always complicated, as the missing rate is typically very high, with low incentives or no incentives at all for companies to report data. The importance of ESG data has also increased dramatically over the last 5 years, resulting in more complete data being released for dates closer in time.

For the selection of stocks, the original dataset included 243 number of stocks. However, the OMXLC has stocks categorized into different tiers/classes, such as Tier A stocks and Tier B stocks, meaning that certain stocks had to be excluded to avoid duplication as they are linked to the same ESG metrics. The Stock that is considered the common one was selected for the analysis. Further, the first year a company was introduced to the market had to be excluded, as no ESG data was available for that year. The data periods also consist of unequal datasets, as a company is evaluated as soon as there is data available for it. In the end, the dataset for 2022 consists of 198 different companies, while the first dataset for 2007 consisted of 105 companies.

4.3 ESG Signal Construction

For my strategy, I have selected a gradient boosting tree (GBT) approach, specifically utilizing the XGBoost library, a powerful tool introduced in 2016 by Tianqi Chen & Carlos Guestrin. GBT is an ensemble learning method that builds a sequence of decision trees, each refining the predictions of its predecessor. An important factor that firstly needs to be addressed is the selection of using a regression or a categorical model. As the object of the model is to closely examine the relationship between ESG metrics and Active Return rather than only categorizing returns into discrete classes, where no understanding of where in the category class in itself placing, Hence, the incorporated model is a regression. Further, XGBoost is particularly apt for working with ESG data, due to its inherent handling of missing values and non-linearity, allowing it to capture complex patterns in the data.

To further investigate the relationship between ESG metrics and Active Return, Shapley Additive exPlanations (SHAP) is incorporated, an approached based on game theory that is used to explain the actual predictions of a model. The inclusion of this approach is based on that by focusing on individual key metrics, noise in the data is excluded and the model can predict more accurately. It is also important to understand in what direction the ESG metrics affect the return, as in if the variable has a positive or a negative effect. The evaluation of the SHAP values is done on the validation period between 2014-2017, where the top 20 ESG metrics are selected and later introduced to create a separate model.

4.4 SHAP value selection

To select the most important features, a Tree explainer utilizing SHAP-values are utilized. 8 different sets of SHAP values were ran on the in-sample period, one for each top 15% of predicted returns and bottom 15% predicted return. Features were selected prioritizing interpretability, meaning that values that are not ambiguous were prioritized and selected. The full list of SHAP values selected can be found in the appendix. The feature list contains 8 Environmental, 8 Governance and 4 Social Metrics. Notably, Salary gap appears consistently across negative effects on all years, while Non-audit to Audit Fees Ratio appears consistently as positive effects. The full list of features and definition can be found in the Appendix.

4.5 Model optimization

The selection of Hyperparameters can vastly change the results of the model and is an important factor in Gradient Boosting Tree (GBT) models. To ensure that the model has the most optimized hyperparameters for the specific task, a Grid search was employed, focusing particularly on optimizing the Mean Squared Error (MSE), specifically focused on optimizing on outliers – top 15 % and bottom 15% of predictions. MSE measures the average squared difference between the observed actual outcomes and the outcomes predicted by the model, indicating model accuracy. Grid Search works by systematically testing a range of hyperparameter values to determine the combination that yields the lowest MSE. It explores a predefined set of hyperparameters, runs the model with each combination of these parameters, and then evaluates the results using MSE as the criterion for performance. While MSE is not a perfect statistical measurement for the task at hand, it is a good indicator of how the model performs. For the model that arrives from utilizing the top 20 ESG metrics utilizing SHAP values, a new grid search was conducted and new Hyperparameters selected.

In the end, the optimized hyperparameters were similar for the 2 models, with a lambda of 10 to overcome overfitting, a moderate learning rate of 0.1 and 50 numbers of estimators, which determines how many trees are tested and averaged on. The Max_Depth for the full model was set to 5, while the Max_Depth for the SHAP was set to 4. Max_Depth indicates how deep the trees are allowed to go.

4.6 Portfolio description and evaluation

In the end, 6 different portfolios are selected with 3 characteristics: Long only (FULL_L, SHAP_L), Short only (FULL_S, SHAP_S), and combined portfolio (FULL_C, SHAP_C). the cut off points to go long in were chosen to be above 20% Active Return and short in below -15% Active Return. These cut-off points ensure a sufficiently large dataset for investment and provide a margin for error in between predicted versus actual Active Return. All the reported portfolios are given an equal weight, which is recalculated each year and weighting depends on how many stocks the model predicts to buy.

To evaluate how the portfolios performed several metrics are reported: Annualized Return(AR), annualized volatility(AV), Sharpe ratios(SR), Alpha, Information ratios(IR), Skewness, kurtosis and number of trades which are used to evaluate the results. Transaction costs are calculated based on the Nordic Nasdaq's standard rate of 0.385 basis points for general investors, and the portfolio is modeled as if it were managed by a U.S.-based manager trading in U.S. dollars. The risk-free rate(RF) is taken from the Ken French library. For statistical check of the model, MSE is reported. Further, a regression analysis is conducted on excess return of the portfolios against the Fama French Five factors – Market (The Benchmark), Size, Value , Profitabilty and Investment. See A Five-Factor Asset Pricing Model' by Fama and French (2014) for further elaboration of the factors.

5. Results & Analysis

5.1 In-sample Results

Exhibit 1: Performance Statistics: 2014-2017

	FULL_L	FULL_S	FULL_C	SHAP_L	SHAP_S	SHAP_C	OMXNLCI
Annualized Return	**21,05%**	-9,50%	-1,86%	**18,25%**	-12,02%	13,11%	10,18%
Annualized Volatility	**20,25%**	20,64%	13,17%	**18,89%**	21,27%	18,74%	16,13%
Sharpe Ratio	**1,03**	-0,47	-0,16	**0,95**	-0,58	0,69	0,62
Kurtosis	**5,94**	9,22	9,57	**5,29**	8,16	7,37	6,68
Max_Drawdown	**-18,30%**	-49,72%	-27,76%	**-19,18%**	-49,63%	-27,17%	-21,79%

Skewness	**-0,17**	0,65	0,4	**-0,12**	0,49	-0,44	-0,42
Average Number of Holdings	**6**	9	15	**5**	9	14	N/A
Annualized Alpha Benchmark	**10,31%**	-0,05%	2,32%	**9,55%**	-3,23%	4,37%	N/A
Information Ratio Benchmark	**0,87**	0	0,21	**0,97**	-0,23	0,45	N/A
Tracking error Benchmark	**0,12**	0,14	0,11	**0,1**	0,14	0,1	N/A
Alpha P-value	**0.107**	0.994	0.686	**0.071**	0.666	0.393	N/A
R-Squared - Benchmark	**0.479**	0.348	0.136	**0.459**	0.311	0.516	N/A

The performance of the FULL portfolios varied significantly as indicated by the exhibit above. The FULL_L portfolio stood out, achieving a notable Alpha of 10.31% against the Benchmark, indicating robust performance. This was despite having a higher AV than OMXNLCI, which was balanced by a superior AR culminating in a SR of 1.03 compared to 0.62 for OMXNLCI. Conversely, FULL_S struggled, recording a -9.50% AR and a high AV of 20.64%. The FULL_C portfolio also underperformed, with a negative AR of -1.86% despite a lower AV, suggesting that only FULL_L managed to surpass the Benchmark. However, the P-value of 0.107 for FULL_L implies that this outperformance might not be statistically significant, indicating that factor beyond the strategy might contribute to FULL_L's superior returns.

Similarly, the Shap Portfolios mirrored the FULL portfolios in their performance patterns. SHAP_L showcased a strong performance with an 18.25% AR and an 18.89% AV, achieving an SR of 0.95 and surpassing the Benchmark. However, the SHAP_L portfolio's P-value was not substantial at 0.071. On the other hand, SHAP_S had a negative AR of -12.02% and an AV of 21.27%, leading to the combined portfolio's underperformance compared to the Benchmark. In terms of metrics, the long portfolios displayed similar performance during the in-sample period, suggesting that selecting the top 20 features might not significantly impact results. However, a closer look at the Mean Squared Error (MSE) of the models reveals that the SHAP_L portfolio is more accurate in its predictions, indicating a stronger performance compared to the FULL portfolios.

Exhibit 2: Regressions against Fama French Five Factors and Mean Squared Error

	In-Sample: 2014-2017						Out-of-Sample: 2018-2022					
	Full_Long	FULL_S	FULL_C	**SHAP_L**	SHAP_S	SHAP_C	**FULL_Long**	FULL_S	FULL_C	**SHAP_L**	SHAP_S	SHAP_C
Alpha (a)	**11,63%**	-0,10%	2,57%	**8,51%**	-2,92%	3,59%	**9,96%**	-1,11%	-3,41%	**16,80%***	-4,97%	12,98%*
	(0,074)	(0,989)	(0,632)	**(0,127)**	(0,701)	(0,488)	**(0.182)**	(0.852)	(0.602)	**(0.024)**	(0.605)	(0.024)
Market (b)	**0,98**	-0,93	-0,43	**0,81**	-0,87	0,9	**1,43**	-1,04	0,55	**1,45**	-0,93	1,37
	(0.000)	(0.000)	(0.003)	**(0.000)**	(0.000)	(0.000)	**(0.000)**	(0.000)	(0.000)	**(0.000)**	(0.000)	(0.000)
Size	**-0,43***	-0,5	-0,54*	**0,03**	-0,45	0,29	**-0,09**	-0,19	-0,03	**-0,04**	-0,66	0,02
	(0.07)	(0.069)	(0.011)	**(0,869)**	(0,124)	(0,135)	**(0.735)**	(0.366)	(0.895)	**(0.887)**	(0.057)	(0.903)
Value	**-0,02**	-0,05	-0,13	**0,03**	-0,26	0,21	**0,07**	-0,70**	-0,46*	**-0,01**	-0,96**	0,13
	(0.953)	(0.866)	(0.590)	**(0,887)**	(0,435)	(0,357)	**(0.741)**	(0.000)	(0.013)	**(0.978)**	(0.001)	(0.407)
Profitability	**-0,76***	-0,55	-0,59	**0,35**	-0,4	0,41	**-0,15**	0,06	0,05	**-0,2**	0	-0,17
	(0,047)	(0.193)	(0.068)	**(0,278)**	(0,373)	(0,184)	**(0.643)**	(0.804)	(0.844)	**(0.527)**	(0.993)	(0.473)
Investment	**0,16**	-0,34	-0,16	**-0,02**	0,01	-0,09	**-0,81***	0,39	-0,08	**-0,50**	0,52	-0,46*
	(0,743)	(0.534)	(0,693)	**(0,970)**	(0,980)	(0,819)	**(0.009)**	(0.109)	(0.763)	**(0.099)**	(0.186)	(0.048)
R^2	**0,54**	0,44	0,34	**0,48**	0,39	0,57	**0,77**	0,77	0,48	**0,77**	0,63	0,83
Information Ratio	**1,04**	-0,01	0,27	**0,88**	-0,22	0,4	**0,68**	-0,09	-0,26	**1,17**	-0,26	1,16
Tracking Error	**0,11**	0,13	0,09	**0,1**	0,13	0,09	**0,15**	0,12	0,13	**0,14**	0,19	0,11

Statistical Measurements - Mean Squared error 2014-2022

	2014	**2015**	**2016**	**2017**	**Average**	**2018**	**2019**	**2020**	**2021**	**2022**	**Average**
FULL - Top 15%	0.06	0.22	0.05	0.11	**0.11**	0.17	0.45	0.82	1.35	0.44	**0.65**
FULL -Bottom 15%	0.07	0.08	0.67	0.07	**0.22**	0.05	0.1	0.07	0.09	0.41	**0.15**
SHAP - Top 15%	0.10	0.13	0.06	0.13	**0.10**	0.08	0.17	0.74	0.43	0.48	**0.38**
SHAP -Bottom 15%	0.09	0.07	0.06	0.08	**0.07**	0.06	0.07	0.13	0.07	0.51	**0.17**

Notes: p-values in (); **,* denote statistical significance at 1 and 5% level, respectively. Alpha, Information Ratio and Tracking error are denoted in annualized terms.

In evaluating the strategy against the FF5, an interesting observation is the increase in alpha to 11.63%, up from 10.31% against the Benchmark for the FULL_L portfolio. Although the P-value also is lower at 0.07 than previously, it is still not below the threshold of 0.05. For the SHAP_L portfolio, the alpha is lower at 8.51%, with a higher p-value than before, indicating that the results are not statistically significant. Generally, none of the Factor's p-values are below 0.05, barring Profitability for the FULL_L and Size for the FULL_C. The negative coefficient for Profitability suggests that the FULL_L portfolio has lower exposure to highly profitable companies or that less profitable companies performed better during the period analyzed. Going forward into the Out-of-sample period, the analysis will focus on the 2 long portfolios, as those held the most promising results. The statistics will however still be reported for the other portfolios, to encapsule if there exist any inconsistencies between the periods.

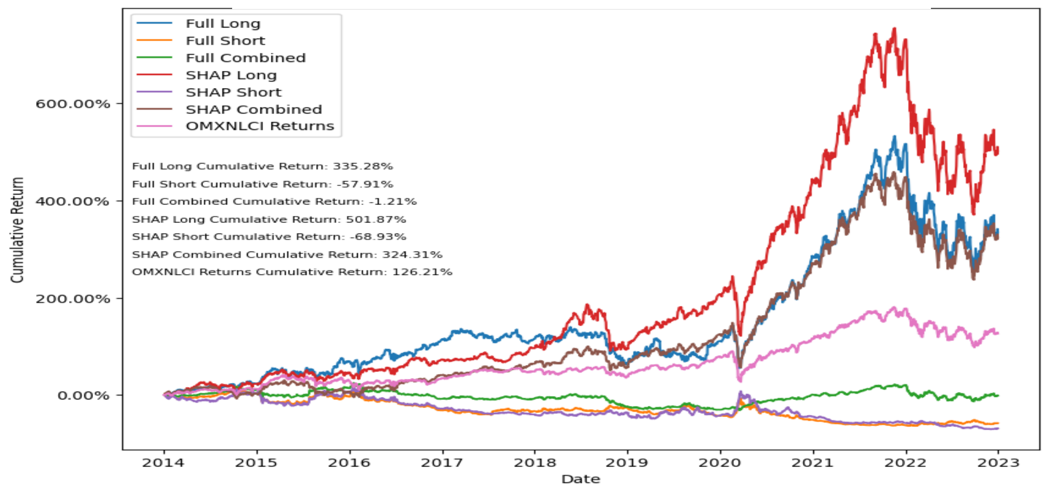
5.2 Out-of-sample results

Similar analysis was conducted on the Out-of-sample results, to see if the results held or if any variability was encountered between the periods. No changes were made from the In-Sample period, but with the prerequisite that the long portfolios are expected to perform well.

Exhibit 3: Performance Statistics: 2018-2022							
Period 2018-2022	**FULL_L**	**FULL_S**	**FULL_C**	**SHAP_L**	**SHAP_S**	**SHAP_C**	**OMXNLCI**
Annualized Return	**14,96%**	-8,79%	1,26%	**24,88%**	-12,21%	20,68%	8,77%
Annualized Volatility	**28,70%**	24,08%	16,69%	**26,83%**	28,59%	24,42%	19,13%
Sharpe Ratio	**0,48**	-0,41	0,01	**0,88**	-0,47	0,8	0,4
Kurtosis	**6,55**	8,92	5,24	**6,35**	5,64	6,49	9,88
Max_Drawdown	**-44,26%**	-61,09%	-32,94%	**-44,97%**	-72,77%	-39,75%	-33,50%
Skewness	**-0,28**	0,62	-0,05	**-0,47**	0,08	-0,52	-0,8
Average Number of Holdings	**18**	10	28	**18**	4	22	N/A
Annualized Alpha Benchmark	**0,05**	0,02	-0,03	**0,13**	0	0,09	N/A
Information Ratio Benchmark	**0,28**	0,16	-0,22	**0,85**	0,01	0,78	N/A
Tracking error Benchmark	**0,16**	0,15	0,15	**0,15**	0,24	0,12	N/A
Alpha P-value Benchmark	**0,55**	0,73	0,63	**0,07**	0,67	0,1	N/A
R-Squared Benchmark	**0,72**	0,63	0,31	**0,74**	0,99	0,81	N/A

For the out-sample period, the results change slightly. While FULL_L performs better than the Benchmark in terms of AR and SR, the performance is substantially lower than the previous period. The SHAP_L portfolio performs similarly to the previous period, but with higher AR (24.88%) and higher AV(26.83%), resulting in an SR of 0.88. The performance is however exceptionally strong in contrast to the Benchmark and the other portfolios, being the only portfolio able to just slightly lower the performance, indicating that the portfolio possibly performs well even in different market conditions. As the model evolves over time with the use of rolling windows, it's reasonable to consider that the improved performance of the SHAP models may be attributed to this factor, as the initial training dataset is quite small.

Exhibit 4 – Cumulative return of portfolios and Benchmark



During the out-of-sample period, the portfolios and the Benchmark experienced notably higher max drawdowns, signalling significant drops from peak values and reflecting a period of higher market volatility. This trend becomes especially pronounced in the graph post 2022, where the Long portfolios and SHAP_combined portfolio exhibit significant fluctuations, reflecting the heightened market volatility triggered by the Ukraine conflict and the onset of interest rate hikes in early 2022. In contrast, the OMXNLCI index displays relatively low volatility, suggesting its resilience and stability during turbulent times, in comparison to the long portfolios and SHAP_Combined portfolio. The short portfolios, while improving compared to the in-sample period, failed to deliver positive returns throughout, suggesting the model's limited success in identifying underperforming stocks based on poor ESG metrics for short-selling.

The SHAP_L portfolio's alpha of 13% indicates substantial outperformance over the Benchmark, yet the p-value of 0.07 lacks statistical significance. Similarly, the FULL_L portfolio, with a p-value of 0.55, also falls short of significance. The increase in the average number of holdings for the Long portfolios is notable, offering diversification and risk mitigation against individual stock failures. However, such diversification within equities doesn't inherently guard against broader market downturns.

In regression against the FF5 factors, the FULL_L portfolio displayed a positive alpha of 9.96%, slightly higher than against the Benchmark alone, but without significant p-value. A significant p-value of 0.009 against the Investment factor, with a Beta of -0.8136, implies inverse performance related to aggressive investment strategies. The SHAP_L portfolio excelled with a statistically significant Alpha of 16.8% (p-value 0.021). Its R-squared value of 0.77 indicates substantial model explanation for returns, though this drops to 0.743 when compared only against the Benchmark, suggesting limited factor explanation. Notably, p-values for most factors exceeded the 0.05 threshold, with the exception of the Size factor for FULL_S and SHAP_S portfolios, and FULL_C at 0.016, implying a better performance likelihood for growth stocks over value stocks.

6. Limitations

The strategy encounters certain limitations, concerning the dataset's limited scope as the analysis is constrained to data available from Refinitiv starting from 2006 when ESG reporting began, thus precluding insights into earlier periods. Given the limitations of the dataset, the back-testing period is also relatively constrained. However, with the use of rolling windows, the model is expected to enhance its performance progressively over time. The strategy is also limited that no currency risks or costs are included in the portfolio. This could possibly have been implemented by assuming a fixed cost for a hedging instrument, but due to the complexity where trading is done in 5 different currencies, it has been excluded from the analysis.

Another limitation is that while focusing on ESG metrics offers a comprehensive view of a company's financial health, it does not encapsulate all factors influencing stock returns. Key elements such as macroeconomic conditions, industry-specific trends, and company financials, which play a role in determining stock performance, fall outside the scope of traditional ESG metrics.

7. Conclusion & Implications

This strategy set out with the primary goal of establishing a link between ESG metrics and Active Return to outperform the OMXNLCI. The analysis involved training two models: a Gradient Boosting Trees (GBT) model using ESG metrics and a second model utilizing SHAP-values to identify the top 20 influential ESG metrics. The results showed that the long portfolios generated significant Alphas when compared to the OMX Nordic Large Cap Index, although the associated p-values were not statistically significant.

Notably, the SHAP_L portfolio emerged as the most effective, surpassing the Benchmark in both Annualized return and Sharpe Ratio. However, the alpha was significant only in the out-sample period when ran against the FF5 factors. This suggests that refining the model and data can enhance performance. However, the significantly High MSE across the periods raises some questions regarding the results. The strong performance of the SHAP_L portfolio in both the in-sample and out-sample periods serves as evidence that emphasizing specific key ESG metrics enhances the model's accuracy, while the stronger performance specifically in the out-sample period highlights the benefits of being trained on more data.

The long portfolios' success in the Nordic stock market can be attributed to the region's focus on sustainable practices and governance, favouring companies with strong ESG profiles. Additionally, the region's trend toward innovation and sustainable operations likely plays a role in these companies' outperformance, positively impacting the long portfolios. A point of concern, not as prevalent in the long portfolios as the others, is the low average number of holdings per period in these portfolios, indicating a high weight per stock and hence increased susceptibility to individual stock movements.

The underperformance of short portfolios in the analysis can be attributable to the strong focus on ESG compliance in the Nordics. Given the region's inclination towards sustainability and ESG-friendly practices, stocks that perform poorly in ESG metrics might already be undervalued or facing market scepticism and the market appears to reward high ESG performers rather than penalizing poor ones. To further establish the validity of the signal created, it needs to be tested on a larger universe such as the Russel 1000 or the s&P 500. By incorporating larger datasets, further analysis can then later be done on a miniscule level, such as comparing stocks from certain sectors against each other. More Hyperparameters could also have been optimized and tested, such as optimizing distinct hyperparameters for the in-sample and out-of-sample periods or by exploring a wider array of parameter combinations.

References

Chen, Tianqi, and Guestrin, Carlos. 2016. "*XGBoost: A Scalable Tree Boosting System*." In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, 785–794. doi:10.1145/2939672.2939785.

De Lucia, Caterina, Pasquale Pazienza, and Mark Bartlett. 2020. *Does Good ESG Lead to Better Financial Performances by Firms? Machine Learning and Logistic Regression Models of Public Enterprises in Europe Sustainability* 12, no. 13: 5317. https://doi.org/10.3390/su12135317

European Securities and Markets Authority (ESMA). *Report on Trends, Risks, and Vulnerabilities*. (2021) URL: https://www.esma.europa.eu/document/esma-report-trends-risks-and-vulnerabilities-no1-2021

Fama, Eugene F., and Kenneth R. French. 2014. *A Five-Factor Asset Pricing Model*. Journal of Financial Economics 116 (1): 1-22. doi:10.1016/j.jfineco.2014.10.010.

Friede, G. (2015, October 22). *ESG and Financial Performance: Aggregated Evidence from More than 2000 Empirical Studies* Sustainable Finance & Investment, Volume 5, Issue 4, p. 210-233, 2015, DOI: 10.1080/20430795.2015.1118917, SSRN: https://ssrn.com/abstract=2699610

Friedman, Jerome H. 1999. "*Greedy Function Approximation: A Gradient Boosting Machine*." Annals of Statistics 29 (5): 1189–1232. doi:10.1214/aos/1013203451.

Jiang, Laurel. *Nordics are Pioneering the Future of ESG*. ESGinvesting, May 2020. URL: https://www.esginvesting.com/2020/05/xx/nordics-are-pioneering-the-future-of-esg/

Lundberg, S. (2017, May 22). *A Unified Approach to Interpreting Model Predictions*. arXiv.org. https://arxiv.org/abs/1705.07874v2

NYU Stern. *Uncovering the Relationship by Aggregating Evidence from 1,000 Plus Studies Published between 2015 – 2020*. New York University Stern School of Business 2021, URL:

https://www.stern.nyu.edu/sites/default/files/assets/documents/ESG%20Paper%20Aug%202021.pdf

Ouyang, Liwen. 2020. "*Seeking Signals from ESG Data*." Bloomberg Enterprise Quants. 2020, URL: https://data.bloomberglp.com/professional/sites/10/798791_SeekingSignalsFromESGData.pdf

Appendix

Exhibit 5: Hyperparameter Exact definition

Hyperparameter	Definition
Lambda	**Lambda**, also known as L2 regularization term, is used to reduce overfitting by penalizing complex models. It adds a penalty for the number and magnitude of the features used, encouraging simpler, more general models.
Learning Rate	**The learning rate**, often denoted as eta, controls how quickly the model fits the residual error using additional base learners. A lower value makes the model training process more conservative, reducing the risk of overfitting but potentially requiring more trees to converge.
Max_Depth	**Max_Depth** sets the maximum depth of a tree in the model. Increasing this value will make the model more complex and capable of capturing finer details, but it also increases the risk of overfitting.
N_estimators	**N_estimators** specify the number of boosting rounds or the number of trees to build. More trees can improve accuracy but also increase computational time and risk of overfitting if too many are used

Exhibit 6: Features selected for SHAP model

SHAP Features	Definition
Total Senior Executives Compensation to Revenues in million	The ratio of total compensation paid to senior executives relative to the company's revenue, indicating how executive pay aligns with company performance.
Net Employment Creation	The net number of jobs created by a company, indicative of its impact on employment and economic contribution.
Policy Board Diversity_True	Indicates whether a company has a policy to promote diversity on its board.
Auditor Tenure	The length of time a company’s auditor has been in position, with longer tenures sometimes raising concerns about auditor independence.
Hazardous Waste	The amount of waste that poses substantial or potential threats to public health or the environment.
Nomination Committee Independence	The degree to which the committee responsible for nominating board members operates independently from company management.
Compensation Committee NonExecutive Members	The presence of non-executive members on the committee that decides executive compensation, which can enhance objectivity.
Announced Layoffs To Total Employees	The ratio of the number of announced layoffs to the total number of employees, indicating workforce stability or restructuring efforts.
Total Hazardous Waste To Revenues USD in million	The ratio of hazardous waste generated to the total revenues of a company, indicating the environmental impact relative to its size.
Executive Compensation Policy	Policies regarding how top executives are compensated, which can include salary, bonuses, stock options, and other benefits.
Average Board Tenure	The average length of time board members have served, which can impact board dynamics and governance stability.
Non-Hazardous Waste	Waste that is not dangerous to the public or environment, often including materials like paper, wood, or certain plastics.
EMS Certified Percent	Percentage of a company’s operations covered under an Environmental Management System, indicating commitment to managing environmental impacts.
Salary Gap	The disparity in compensation between different levels within a company, often examined in the context of pay equity.
GHG Emissions Indirect, Scope 2 to Revenue USD in million	Greenhouse gas emissions produced indirectly (e.g., from purchased electricity) relative to the company's revenue.
Human Rights Breaches Contractor_True	Indicates whether there have been any human rights violations by contractors associated with the company, reflecting on its ethical supply chain management.
Profit Warnings	Announcements made when a company expects that its profits will not meet analyst expectations, which can be an indicator of governance and financial management quality.
Energy Purchased Direct	The total amount of energy directly purchased by the company, indicating its direct energy footprint.
Water Recycled	The volume of water recycled by the company, reflecting its efficiency in water use and conservation efforts.
Energy Purchased Direct	The total amount of energy directly purchased by the company, indicating its direct energy footprint.

Exhibit 7: Graph of missing rate in ESG Metrics.

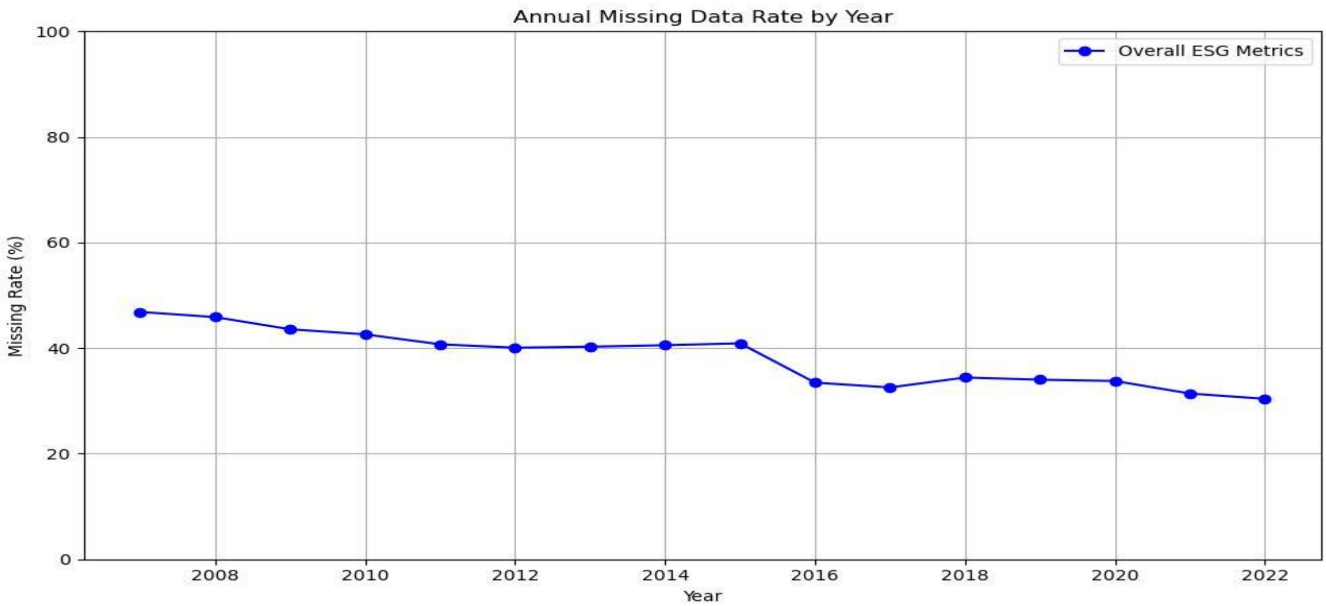
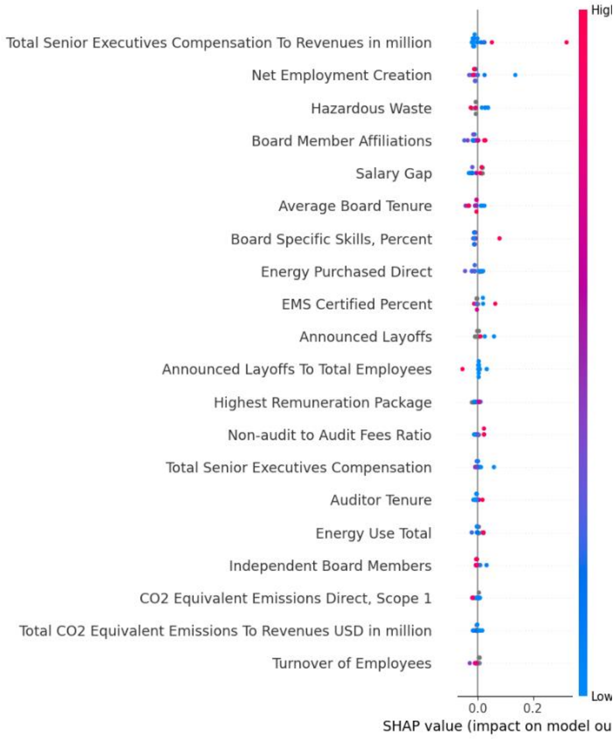
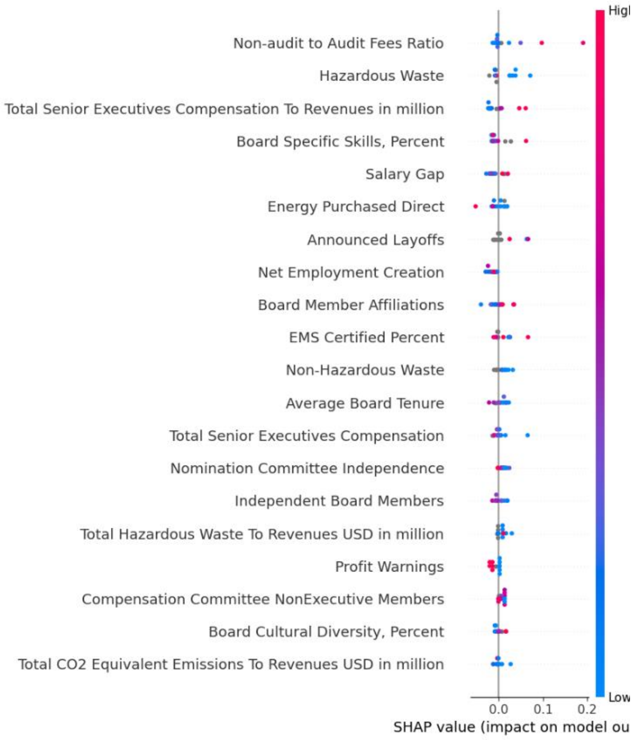


Exhibit 8: SHAP Plots: 2014

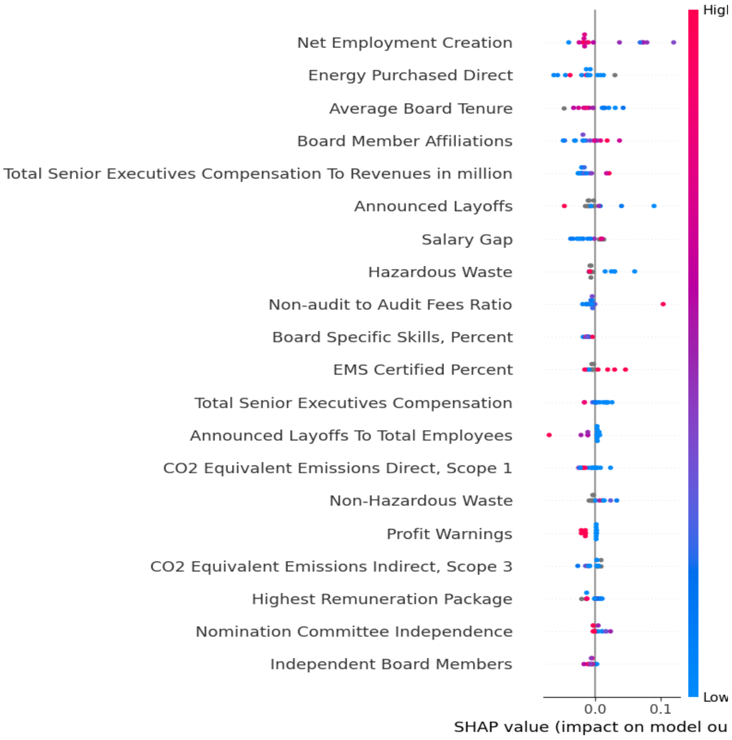


Bottom 15%- 2014



Top 15 %- 2014

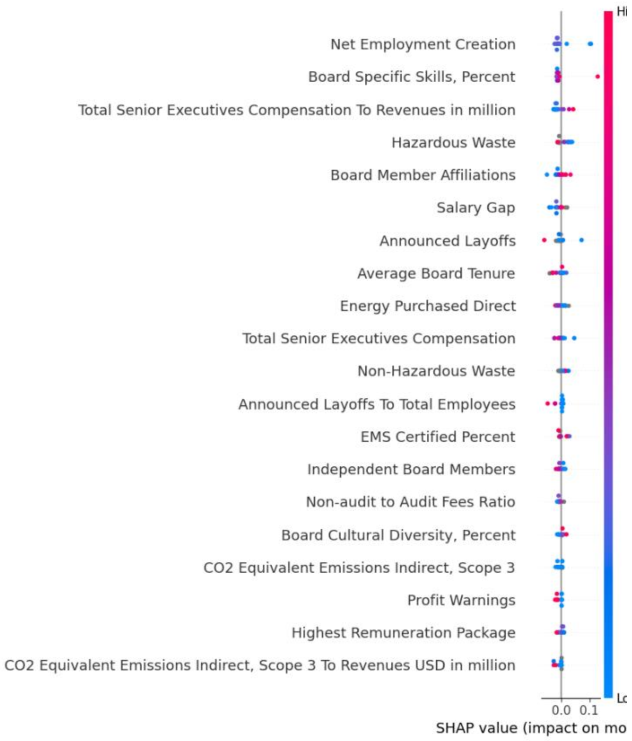
Exhibit 9: SHAP Plots: 2015



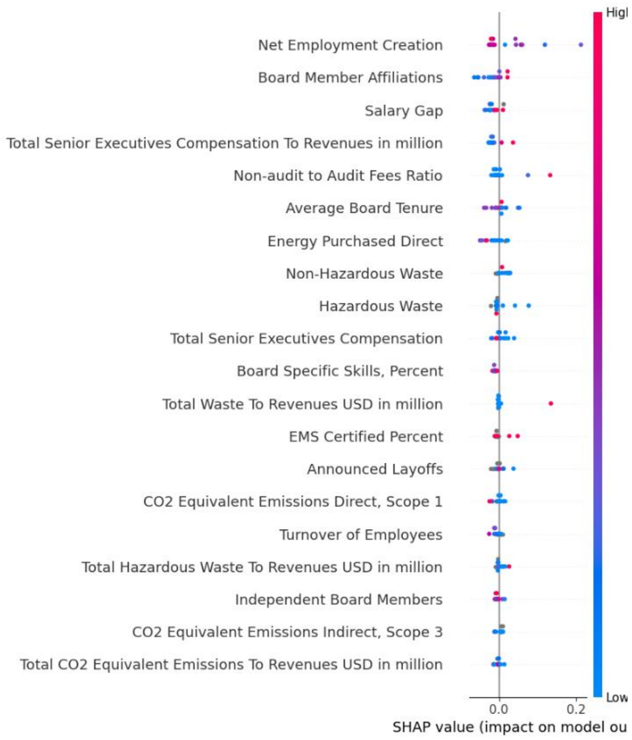
Bottom 15 % -2015

Top 15 % - 2015

Exhibit 10: SHAP Plots: 2016

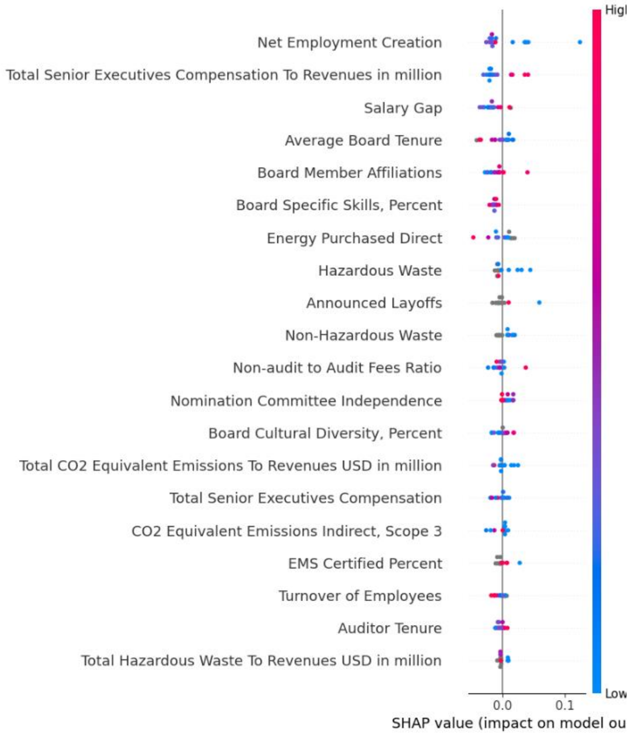


Bottom 15 % - 2016

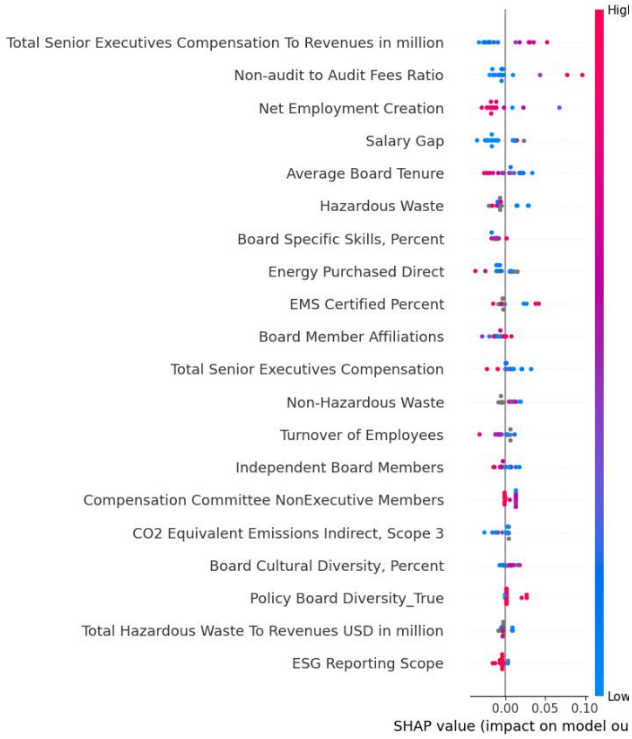


Top 15 % - 2016

Exhibit 11: SHAP Plots: 2017



Bottom 15 % - 2017



Top 15 % - 2017