Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Economics and Management
Department of Economics

Thomas Vincent Borgersen

ESG and Machine Learning in Asset Pricing

Master's thesis in Financial Economics
Supervisor: Ranik Raaen Wahlstrøm
June 2022

NTNU
Kunnskap for en bedre verden

Thomas Vincent Borgersen

ESG and Machine Learning in Asset Pricing

Master's thesis in Financial Economics
Supervisor: Ranik Raaen Wahlstrøm
June 2022

Norwegian University of Science and Technology
Faculty of Economics and Management
Department of Economics

NTNU
Norwegian University of
Science and Technology

Preface

I would like to thank my supervisor, Ranik Raaen Wahlstrøm for providing exceptional guidance through this process. I am particularly grateful for the valuable insights in coding-related processes, which I hope to be of great utility for myself and my peers in handling complex problems in my upcoming professional career.

Abstract

This thesis utilizes machine learning methods to identify and test suitable variables for explaining the cross section of stock returns. Further, it assesses the value of including non-financial variables derived from firms' Environmental, Social and Governance (ESG) reporting. In the analyses, both explanatory power, model parsimony, and overfitting are considered.

First, a broad set of variables suggested in the existing literature is identified. These are selected based on the XGBoost (Extreme Gradient Boosting) and Lasso (Least Absolute Shrinkage and Selection Operator) frameworks. The resulting set of variables is shown to outperform the original set with respect to goodness-of-fit criteria, also when accounting for model parsimony. Second, this thesis conducts a comparison of two variable sets where variables based on ESG reporting are included and excluded, respectively. The inclusion of the ESG variables is found to insufficiently increase the goodness-of-fit criteria, as well as decreasing it when accounting for model parsimony. This suggests that the ESG variables offer little valuable information to investors. Finally, an out-of-sample analysis reveals overall low explanatory power for both ESG and non-ESG variables. This implies overfitting and that the variables are not suitable for making predictions.

The accuracy of findings in this thesis might be seriously compromised by a variety of reasons, most prominently data related issues causing omitted variable bias. This confirms suggestions in previous literature that the lack of sufficient ESG reporting impedes investors from incorporating it in investment decisions. However, missing values are likely to be less prominent in the future. In addition, certain adaptions can be made to the research methodology to better cope with the current ones. Including but not limited to improving data quality by sourcing it from more databases.

Sammendrag

I denne masteroppgaven brukes diverse maskinlæringsmetoder til å identifisere, samt teste, passende variabler for å forklare aksjeavkastningen i et tverrsnitt. Videre vurderes verdien av å inkludere ikke-finansielle variabler fra rapporterte miljø-, sosiale- og forretningsetiske forhold. I analysen vektlegges bade forklaringskraft, samt antallet forklaringsvariabler og mulig overtilpasning.

Først ble et bredt spekter av forklaringsvariabler, hentet fra relevant litteratur, automatisk filtrert ved hjelp av XGBoost (Extreme Gradient Boosting) og Lasso (Least Absolute Shrinkage and Selection Operator). Det resulterende variabelsettet viste seg å utkonkurrere det opprinnelige, både basert på ren forklaringskraft og når antallet forklaringsvariabler ble hensyntatt. De samme kriteriene ble også brukt til å sammenligne to datasett, hvor et inneholdt ESG-variabler og det andre ikke. ESG variablene viste seg å utilstrekkelig øke forklaringskraften, og til og med senke den når antallet ekstra variabler ble hensyntatt. Dette antyder at ESG-variablene tilfører lite nyttig informasjon til investorer. Når modellene ble testet på usett data, resulterte dette i svært lav forklaringskraft, noe som antyder at de er overtilpasset og egner seg dårlig for prediksjon.

Det er flere grunner til å tro at resultatene i denne masteroppgaven har svekket troverdighet. Først og fremst grunnet datarelaterte problemer som resulterer i «utelatt variabel problem». Dette er i tråd med tidligere studier som også peker på lav datakvalitet som noe av det som hindrer investorer i å bruke ESG-data i investeringsbeslutninger. Det virker derimot til at disse hindringene vil bli mindre i fremtiden, i tillegg til at det finnes en rekke konkrete tiltak for å utvikle forskningsmetoden for å bedre håndtere dagens utfordringer. Eksempelvis å heve kvaliteten på datasettet ved å benytte flere databaser.

Contents

1 Introduction

Most of global consumption is realized trough the interaction with some type of market. Being a local grocery store, an online clothing store, or trading platform for financial instruments. As well known from basic economic market theory as explained in Riis and Moen (2017), prices and volumes in such markets are decided by the equilibrium point. In a perfect market this point maximizes the utility of both consumers and producers, and thus also the socioeconomic utility.

Furthermore, an effect from economic activity on a third party that is outside or *"external"* to the market, is called an externality. This external cost or benefit is not reflected in the prices of goods and services consumed, but may affect the level of production or consumption that would optimize utility for the broad society. When having a beneficial affect on the third party it is reffered to as a positive externality, and a negative externality in the opposite case. An example of the former would be the additional technology or knowledge available to society as a consequence of investment in innovation and education. A timely example of the latter is $CO_{2,}$ or other green house gases, emitted during the production and consumption of fossil fuels. As the destructive effects on the shared resource that is our planet, is arguably not reflected in the costs of production and consumption of CO_2-intensive goods and services.

Furthermore, the heavily debated *"Shareholder theory"* or *"Friedman doctrine"* is a normative theory originating from Friedman (1962). He controversially stated that a corporation's only social responsibility is to please its shareholders (Calhoun, 2017). Moreover, the total value of an enterprise can be calculated directly by estimating the present value of of debt and equity. Where the value of the equity can be calculated directly by estimating the present value of the future cash flows going to shareholdres, which is done in the Dividend Discount Model. An alternative method is calculating the present value cash flows from a firms investments and operations, and thus subtract the present value of debt calims on these cash flow. This is done in the Discounted Cash Flow Model (Penman, 2013). Evidently, if shareholders would only care about the returns of their portfolios, their only concern would be the cash flows produced by the underlying asset of the stocks they own. Thus, if the company were to be responsible for negative externalities, this would not be of concern of the investor as long as this *"price"* is payed by the broad society and not by the company, directly or inderectly. Consequently, regulators routinely put a price on externalities, either trough subsidies or taxes, or by forcing change of production volumes. Ideally, both methods have the same effect of pushing the market equilibrium of prices and volumes to a point that maximizes socioeconomic utility (Riis and Moen, 2017).

Therefore, in recent years, nation states and other regulatory institutions have increasingly regulated various markets in the interest of the broader society (Divanbeigi and Ramalho, 2015). This could be regulations on production processes or how emissions and waste should be handled (EPA, 2021). Such regulations could have significant financial implications on the companies compelled to change processes to comply with regulations, or in the form of fines or other penalties if failing to do so.

As new regulations might have implications for the companies' financials, this is clearly of concern of investors (Zeidan, 2021). The growing interest in non-financial reporting in general emerges from the increased focus on climate change and ecological issues. It is also amplified by a variety of public scandals from large companies during the past decades. One of the more memorable being the BP Gulf of Mexico oil spills in 2010 (EPA, 2022a). Resulting in weakened stakeholder trust and increased focus on gaining better and more transparent information from companies (Amran and Ooi, 2014; Uyar, 2016; Aluchna and Roszkowska-Menkes, 2019).

Hence the focus on the Economic, Social and Governance (ESG), often referred to as *"sustainability"*, has increased dramatically, for both investors, companies, politicians, and the broader society. Consequently, the perceived value of a company is increasingly dependent on to what degree investors deem a company to be compliant with current regulations. Further, the ability of investors to assess the value of a company is not limited to current regulations and market trends, but also the impacts of future ones, as this would have an implication on future earnings and risks, and therefore the present value used for value evaluation (Penman, 2013). Perhaps the most prominent speculations in this regard have been on the implications of climate change mitigating regulations and trends. Investors pour tremendous amounts of money into companies with stated missions of contributing to a sustainable future, either through the production of cleaner or less carbon intensive energy productions, such as renewable energy, or even taking CO_2 out of the atmosphere trough carbon capture and storage (McKinsey & Company, 2018; The Economist, 2021).

Consequently, investors have progressively attempted to incorporate these ESG factors in the investment decisions when constructing portfolios. According to Morningstar (2022) nearly $70 Billion was poured into *"sustainable funds"*. Further, (UNPRI, 2020) estimates that as much as $103,4 trillion worth of assets was under management in March 2020 with a mandate including some degree of responsible investing. These mandates can incorporate activities such as excluding companies based on certain criteria as done by the Norway's Sovereign Wealth fund, based on considerations of ethical and environmental concerns (NBIM, 2021a). Another example of such mandates is that of the mutual fund Storebrand Global ESG where 95% of its investments mimics the MSCI World Net Index, while 5% is invested based on the strategy of including companies with high sustainability rating and excluding those with low ones (Storebrand, 2022).

During 2020-2021, the number of companies gone public have skyrocketed (EY, 2021a). Many of these have received extreme valuations based on mere plans to develop some new technology that will drive the world towards a more sustainable path. As the market seems to reward *"green"* activities, companies have increasingly marketed these activities in terms of moving them into separate companies which in turn is offered to investors as new "purely green" companies. (EY, 2021b; Financial Times, 2021). However, is it only the revolutionary *"green"* technology companies that are rewarded for their green activities, or is more incremental improvements towards more sustainable business also rewarded in the form of higher asset prices and larger returns for investors?

There have been a large variety of proposed asset pricing models over the years. One of the most widely known is the single factor Capital Asset Pricing Model (CAPM) (Sharpe, 1964; Lintner, 1965; Mossin, 1966), which explains the returns and prices of financial assets based on one factor – systematic risk represented by the Beta. Later, Fama *et al.* (1992) introduced the three-factor model to better explain assets' returns and prices, introducing a firms size, measured by market capitalization, and it's book-to-market ratio as additional risk factors After this, a variety of factors based on information on ESG and responsible investing have been introduced for existing asset pricing models in order to improve them. In the process of empirically finding new variables, however, researchers might fall into the trap of "data mining" or *"data snooping"* (Black, 1992).

In this thesis, factors based on ESG models for use in asset pricing models are analyzed. To avoid narrow assumptions from previous research, a broad set of variables is included in the dataset. However, since standard linear regression models, such as the OLS have been showed to not deal with high dimensionality and collinearity among variables, various machine learning methods to handle the problems of high dimensionality and collinearity are utilized (Chen, Pelger and Zhu, 2019; Bryzgalova, Pelger and Zhu, 2020).

The following Research Questions (RQs) will thus be investigated:

RQ1: Can machine learning methods such as Lasso- and XGBoost regressions be utilized to identify suitable factors in asset pricing models for explaining the cross section of stock returns for companies listed in the S&P 500 during the last 10 years?

RQ2: How important are factors measuring ecological and climate change when explaining the cross section of stock returns for companies listed in the S&P 500 during the last 10 years?

RQ3: What implications does the findings in this thesis have for the future potential and returns of ESG investing strategies?

To answer these questions, several machine learning regression methods is applied on a cross-sectional dataset of firm characteristics, macroeconomic factors, ESG-related variables, and the total returns on all individual stocks listed in the S&P 500 from 2012-2021. The broad variety of variables is selected based on suggestions from existing literature and used to partial out the effects not attributable to the ESG-variables. To include only the most relevant variables, Lasso and XGBoost is used for automat variable selection. Filtering out the variables with poor explanatory power and reduces dimensionality to a level suitable for Ordinary Least Squares (OLS) regression. The performance of the resulting set of variables is then compared to the initial one, based on goodness-of-fit criteria such as R-squared (R^2) and mean squared errors (MSE), also when accounting for model parsimony measured by Adjusted R^2, Bayes Information Criterion (BIC) and Akaike's Information Criterion (AIC). These same metrics are then be used to compare the relative performance of two datasets where ESG-variables are included and excluded, respectively. Furthermore, regressions are then performed on unseen data to assess whether the models contain good out-of-sample predictability. The relative importance and economic implications of the most important variables from the second-generation variable selection is then determined. This by analyzing the coefficients and associated p-values for OLS, as well as the SHAP-plot from the XGBoost regressions.

The automatic variable selection method was proven to be effective, as the resulting set of variables outperformed the initial one. Moreover, from performing the different regression methods on the various data subsets, the ESG variables were shown to add some explanatory power to the models. However, when considering the various penalizing criteria, the models containing ESG variables was considered less favorable than the ones where they were excluded. The scores were, on the other hand, did vary among the various penalizing criteria and across time. Where the most parsimonious models were most frequently favored by BIC and least frequently by Adjusted R^2. Furthermore, the R^2 for out-of-sample regressions was very low, and even negative in certain time windows, for both the OLS and XGBoost. This implies that the relationships between explanatory variables and total returns are unlikely to be consistent with the true relationships.

The most impactful explanatory variables were mainly ones shown to be so in a majority of existing literature. However, the coefficients and SHAP-values from several of these variables suggested impacts on predictions that was inconsistent with literature and bared counterintuitive economic implications. Furthermore, none of the ESG-variables were deemed significant by OLS model. On the other hand, CO_2 emissions-to-reveneue, direct CO_2 emissions, water-use-to-revenue and energy use was included as one of the 20 most important variables in several of the SHAP-plots. This diversity may be caused by XGBoost identifies relationships between these ESG variables and returns, not picked up on by OLS due to the lack of flexibility. The impacts from all ESG variables on expected total returns is ambiguous and seem to imply that the predicted returns are marginally smaller for more favorable ESG-values, all else equal.

The implications of the findings in this thesis are that ESG variables do not have a particularly strong impact on stock valuations or investment decisions. Moreover, the ambiguity and slight negative impacts on predictions, may also support existing literature. Suggesting that ESG-investing strategies might outperform the broad market in the short term due to capital inflow in Green assets but underperform in the long run once equilibrium is reached.

However, the empirical findings are likely to be compromised, predominantly due to data related issues. One is the large number of missing observations, particularly for the ESG variables. All missing observations was filled in, using K-Nearest Neighbors (KNN) imputation, but the quality of the resulting dataset is questionable. The fraction of missing observations is further unevenly distributed over time, with far fewer missing observations in recent years. This means that observations from newer years is overrepresented, which might cause a *“selection bias”*. Moreover, data measurement on annual frequency limit the number of observation for time-fixed variables to ten in total, five for each time window. This strongly contributes to overfitting and poor out-of-sample performance. Furthermore, *“omitted variable bias”* is likely to be present by several reasons. One is that the several lagged or differenced variables is not included. Moreover, the models do not account for different sensitivities across different companies or industries, which likely to cause *“heterogeneity bias”*. Finally, multicollinearity still present between the explanatory variables might cause the estimated coefficients and p-values for the OLS to be unreliable. The latter is on the contrary not assumed to compromise the results from XGBoost and Lasso. Finally, several concrete changes to the research methodology that is likely enabling it to better cope with the mentioned issues are identified.

The rest of this thesis is structured as follows: In *Section 2*, relevant literature on traditional asset pricing methods is reviewed, with additional subsections concerning the role of ESG in such frameworks, and the potential for improving them through utilizing machine learning. In *Section 3* the data is presented. *Section 4* presents the estimation methods as well as preprocessing the data. *Section 5* presents and compares the empirical results, as well a discussion around their economic implications. *Section 6* concludes the thesis with a summary and recommendations for further work.

2 Literature review

2.1 Asset pricing

As mentioned in the introduction, fundamental analysis as explained in Penman (2013) is an attempt to value a stock based on the present value of the payments the owner of the stock will receive. These payments are predicted by analyzing the anticipated earnings and dividends of a company, as well as expected future interest rates and risk evaluation. An analysis of a firm's anticipated future performance can suggest that the intrinsic value of the firm is higher or lower than the current market value of the company. A higher intrinsic value would suggest that the stock is undervalued by the market, and the investor could therefore profit by purchasing the stock. Otherwise, if the intrinsic value is lower than the market value, the stock is overvalued, and an investor can profit by selling it short. Finally, the efficient market hypothesis as explained in Bodie, Kane and Marcus (2020a), suggests that markets are efficient. That is, all investors have the same information and therefore all the available information about future performance is already reflected in the stock's price. In this case, there is no possibilities for profiting by buying or selling any stock.

Also mentioned in the introduction is the method of estimating the intrinsic value of the company is through the Discounted Cashflow Model. However, the value of future cash flows is not certain. Consequently, the perceived value of a company, and thus its stock price, would change in accordance to changes in assumptions about the future cash flows, such as perceived future company profits and discount rate changes over time. As the returns are unknown, investors need to account for the uncertainty of the returns of the assets by spreading their bets. This section next describes how Modern Portfolio Theory, the Capital Asset Pricing Model and Arbitrage Price Theory explains how investors deal with uncertainties of returns. Finally, this section addresses how these theories are extended to include non-financial factors when considering the role of corporate responsibility.

2.1.1 Modern portfolio theory

Markowitz (1952) introduced the *"modern portfolio theory"* (MPT) and the mean-variance model. This framework makes four assumptions. First, it assumes that investors cannot know for certain which assets will yield the highest return over a given period. The investors can, however, calculate the probability distribution for each asset trough finding the expected value of return and the associated variance of this return.

Second, this framework that all investors seek to maximize the level of return that can be obtained for each level of risk. Sharpe (1966) expressed the level of excess return for each level of risk as a ratio of excess return to volatility in what has been famously known as the Sharpe-ratio:

$$S_p = \frac{E(r_p) - r_f}{\sigma_p} \tag{1}$$

where r_f is the risk-free rate, and $E(r_p)$ and σ_p is the expected rate of return and the standard deviation of a portfolio p, respectively.

Third, it is assumed that the total uncertainty of a stock's return can be divided into systematic and unsystematic risk factors. Unsystematic, or firm-specific risk factors, can be neutralized trough owning a variety of different firms, also known as diversification. Hence, the only type of risk that ads to the risk of the total portfolio is systematic risk. In other words, it is only the covariance term between each asset and the portfolio that ads to the total risk of the portfolio. Thus, for each level of return, a minimum variance portfolio can be obtained through diversification. This can be plotted in a return-risk diagram as the minimum variance frontier as shown in *Figure 2-1*. The part of the minimum variance frontier that is above the global minimum variance is not pareto dominated by any portfolio, in that no portfolio can yield both higher return and lower variance at the same time, and this part is called the efficient frontier.

Hence Markowitz (1952) explained that the expected returns and a covariance matrix, along with the assumption that all weights of stocks measured as a fraction of the total portfolio value is used to maximize the Sharpe ratio to obtain the market portfolio. Showed schematically in *Figure 2-1*.

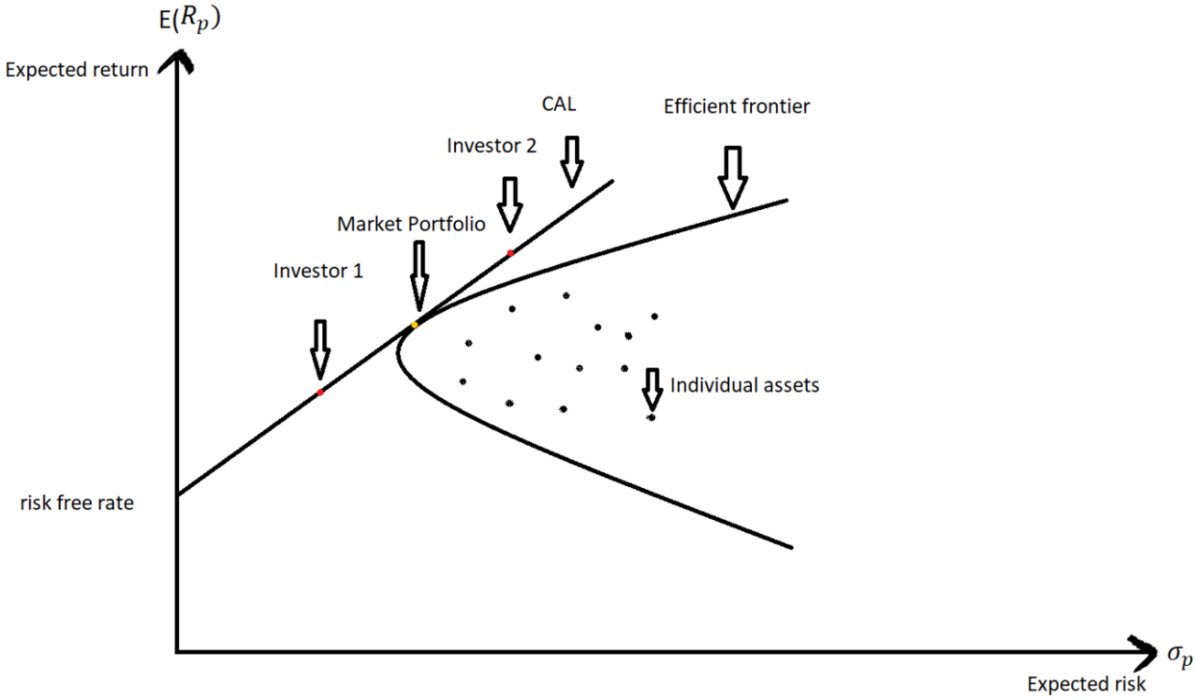

Figure 2-1. The figure illustrates a plot of the efficient frontier which minimizes the risk for all levels of expected returns. The market portfolio is indicated as the tangent point between the CAL and the efficient frontier. It additionally illustrates how two investors can end up on different points on the CAL due to individual levels of risk aversion.

Fourth, it is assumed that different investors have different levels of risk aversion. Meaning that individual investors may gain different levels of utility for different levels of risk versus return. Moreover, government bonds such as the three-month US treasury bills are assumed to have virtually no risk associated with them. Hence the rate of return one would receive from these *"risk-free"* assets are used to approximate a risk-free rate of return. As each investor would choose the same risky portfolio, being a share of the value weighted market portfolio, individual preferences for risk would only be expressed through each individual investor's capital allocation between the risky portfolio and the risk-free asset. This creates different individual complete portfolios with different levels of risk and expected returns, which would be located at different points on the tangent line. The tangent line is therefore called the Capital Allocation Line (CAL). This is expressed in *Figure 2-1*, where investor 1 has a smaller expected risk and return because of having a larger portion invested in the risk-free asset than investor 2. As investor 2 has a higher expected return and risk than the market portfolio, this shows that investor 2 actually has a short position, meaning borrowing at the risk-free rate, and allocate all assets and borrowings in the risky market portfolio.

Previously, the only constraints for finding the efficient frontier of portfolios were that weights had to sum to 1, in addition to optimizing for a fixed level of either volatility or expected return. Then, step 2 is finding the point of the tangency portfolio, which is the portfolio that maximizes the Sharpe-ratio. Furthermore, a client can have individual constraints such as minimum dividend yield, no short positions, or ethical considerations such as ESG-aspects. These additional constraints will then result in a different efficient frontier, where the tangency portfolio determined by the efficient frontier and the risk-free rate result in a less steep CAL and thus Sharpe-ratio (Bodie, Kane and Marcus, 2020b).

Furthermore, Bodie, Kane and Marcus (2020b) argue that even though the MPT provides a simple as well as intuitive approach to portfolio optimization, the required input list is quite difficult to obtain in real life. Both as the expected returns of companies are hard to identify, but particularly the correlation matrix is hard to obtain.

2.1.2 Single factor models

The theory of obtaining the optimal portfolio as laid out by Markowitz focuses on how each individual investor should allocate wealth based on the existing prices in the market. The CAPM explains how these market prices are obtained in equilibrium. Models that explain expected return as a risk premium to the underlying risk factors are generally called *"factor models"*. As explained in the MPT, the factor models distinguish between systematic risk and unsystematic risk. One of the most simple and famous factor models is the CAPM, which attributes all the systematic risk into a single market risk factor. The asset's price sensitivity against this market risk factor is denoted as the beta (β), which says how much the price change of the asset correlates with the price change of the market portfolio.

As all investors optimize their portfolios using Markovitz model for efficient diversification, one asset having a different risk-return relationship than the market portfolio would incentivize investors to rebalance their portfolio until it had the risk-return relationship as the market portfolio, mathematically expressed as:

$$\frac{E(R_M)}{\sigma_M^2} = \frac{E(R_i)}{Cov(R_i, R_M)} \tag{2}$$

where $E(R_M) = E(R_m - r_f)$ is the expected excess return, being the difference between the expected return of the market portfolio R_m and the risk-free rate r_f. σ_M^2 is the variance of the market portfolio, $E(R_i)$ is the expected excess return to company i, and $Cov(R_i, R_M)$ is the covariance between stock i and the market portfolio. Equation (2) can be rewritten to

$$\beta_i = \frac{Cov(R_i, R_M)}{\sigma_M^2} \tag{3}$$

expressing the additional variance to the market portfolio stemming from a single asset i, as a fraction of the variance of the Market portfolio. Finally, reaching the expression of the expected return $E(R_i)$ of a financial asset i as a function of its β_i:

$$E(R_i) \; = \; \beta_i[E(R_m - r_f)] \tag{4}$$

This equation represents one of the most prominent predictions of the CAPM, that the expected excess return of an asset is determined by the sum of the risk-free rate and the market risk premium proportional to the asset's beta (β_i). Finally, the firm specific size of this premium is determined only by the firm's beta, without considering the total variance of the firm. Hence $\beta_i[E(R_m - r_f)]$ represents the expected risk premium of asset *i.*

Another important theoretical framework for deriving asset pricing models is *"the arbitrage pricing theory"* (APT), presented by (Ross, 1976). As described for the MPT, the APT comparably predicts asset prices by deriving models linking risks and returns. The two theories diverge however, in the way the models are derived.

APT assumes that investors would trade as large volumes as possible to maximize the gains from an arbitrage opportunity, thus drive the price of the cheap asset up and the expensive one down until it would reach equilibrium and the arbitrage opportunity eliminated. The MPT, on the other hand, would reach the same conclusion, but with an alternative explanation. Namely, that all investors rebalance their identical portfolios by increasing the weight of the underpriced asset and decreasing the weight of the overpriced asset. It is finally the sum of the small changes in many portfolios that drives the prices to a new equilibrium, opposed to a few investors trading large amounts in the APT.

2.1.3 Multiple factor models

Assuming APT, asset pricing models such as the CAPM can be extended to include multiple factors of systemic risks, and the excess return can be expressed as a multifactor model, done by Rosenberg and Guy, (1995):

$$R_i = E(R_i) + \beta_{i1}F_1 + \beta_{i2}F_2 + \cdots + \beta_{ip}F_p + e_i \tag{5}$$

where R_i is the excess return of asset i, $F_1, F_2, \dots, F_p$ are the various macro factors of risk, $\beta_{i1}, \beta_{i2}, \dots, \beta_{iK}$ are the sensitivities to each specific macro risk factor, and e_i is the firm-specific or non-systematic risk. All macro risk factors have an expected value of zero, meaning that non-zero values would imply surprises to the market. Hence, the excess return of the asset will equal it's expected excess return plus the respective *"surprise"* times the individual sensitivities for each of the macro risk factors, in addition to the surprise of firm-specific influence on the returns. One way to identify the most likely sources of systemic risk is from using a multifactor CAPM (Bodie, Kane and Marcus, 2020b), where these additional factors of systemic risk represent investors demands to hedge for risks associated with investment opportunities or consumption. However, the dominating approach in recent years have been to empirically identify the firm characteristics doing a good job of explaining returns. The factors are chosen based on their ability to explain past average returns, implying that they capture risk premiums. The three-factor model of Fama and French (1996) is an example of the latter and have come to which have come to be one of the commanding method in empirical research of security returns. Mathematically expressed:

$$R_i = \beta_{iM} R_{iM} + \beta_{iSMB} SMB_i + \beta_{iHML} HML_i + e_i \tag{6}$$

where SMB (Small Minus Big) represents the return of a portfolio consisting of small stocks in excess of a portfolio of large stocks. Similarly, the HML (High Minus Low) represent the returns of a portfolio of stocks with high book-to-market ratios in excess of one with low book-to-market ratios. M is the market index, capturing systemic risk from macroeconomic factors. SMB and HML is then supposed to capture the deviations from the returns predicted by the CAPM.

The several variants of the three-factor model introduced in Fama and French (1993) are currently the leading multifactor models. In addition to the factors themselves, Fama and French also introduced a general method of generating factor portfolios and suggested that adding an extra factor to the model, for example size, would suggest a higher expected return for small firms than predicted by the single-single factor CAPM. Having a higher expected return than the CAPM predicts, is often referred to as having a positive alpha (α).

However, it is indicated that smaller firms are typically worse of in recessions, which would motivate investors to shy away when these are expected and would explain the risk premium of smaller firms. What appears to be an alpha in this model could be viewed as an additional structural risk-factor in a multifactor model. However it is not clear if the alpha values achieved trough utilizing the multifactor models actually represent mispricing of an asset, or if the additional firm characteristics actually represent additional risk factors that are hard to identify (Bodie, Kane and Marcus, 2020b).

Furthermore Liew and Vassalou (2000) test whether high-book-to market ratio (HML), size (SMB) and momentum (WML) can be linked to future Gross Domestic Product (GDP). They find that GDP seem to be predicted by HML and SMB, but little evidence that this is the case for WML. The intuition behind why this is the case with HML and SMB is that they proxy business cycle risk. Petkova and Zhang (2005) identifies that the higher returns for companies with high book-to-market ratios is due to increased risk in periods of lower economic growth. This because of larger amount of tangible capital, which would make them suffer from excess capacity in recessions. On the contrary, growth firms with typically less tangible capital and hence lower book-to-market ratios can solve this problem by postponing or shelving investment plans.

Jegadeesh and Titman (1993) found that stocks that previously had performed well continued to do so for three to twelve months in the future, and vice versa for stocks that preciously performed poorly. This phenomenon is termed *"momentum"* and was included by Carhart (1997) as a fourth factor to the standard three factor model of Fama and French, as Winners Minus Losers (WML), demonstrating that the alpha of mutual funds was not attributed by clever stock picking by investors, but merely to mutual funds tendency to be highly weighted in stocks that previously had performed well. It is not however straight forward to explain this in terms of a proxy for a risk factor, and this phenomenon is hence often explained by behavioral economics. Chan, Karceski and Lakonishok (2003) found that investors overestimated earning growth rates of firms that previously had performed well, and therefore overpriced the value of these firms.

However, Black (1992) points out the problem of *"data mining"* or *"data snooping"*. This is when a researcher performs a study in a range of different ways, with different explanatory variables in different periods and with different models. If only the results that support the researcher's conclusion are reported, the results might seem significant, but in reality, only accidental. This can be mitigated by reporting all the runs of the study, even results that contradict the conclusion. Black further points to the fact that it is usually only the most striking results that wind up being published. Furthermore, researcher tends to build upon each other's works, in terms of often utilizing roughly the same variables and even datasets. Which makes it likely that even results from different researchers are results of datamining, as one results build on the *"blind start and false alley"* of another.

2.2 ESG in asset pricing

Friede, Busch and Bassen (2015) reviews over 2000 studies on the effect ESG-related matters have on financial performance. They found a positive relationship between ESG and financial performance in a large majority of the studies, and a non-negative relationship in over 90% of the studies. Lins *et al.* (2017) found that firms with high scores of Corporate Social Responsibility (CSR) outperformed firms with lower CSR-scores during the same period. This suggests that firms with higher CSR-grades was less risky than the firms of low CSR-scores. Both suggesting a solid case for *"ESG-investing"*.

On the contrary, Hong and Kacperczyk (2009) suggest that stocks that are generally viewed as *"sin-full",* such as producers of tobacco and alcohol, are frequently left out of the portfolios of the institutions that are bound by societal norms. The authors conclude that these institutions and other players that do not invest in *"sin-stocks"* by the same reason, in fact pay a financial cost by not being invested in these stocks. It is further pointed out that these types of stocks are less frequently held by mutual funds and hedge funds, even though the *"sin-stocks"* could help these funds fulfill their mandates of mitigating risks by diversifying their portfolios. Additionally, *"sin-stocks"* receives less media-coverage than other stocks, making the price less likely to be trending and overbought by especially amateur investors. Moreover, Statman, Fisher and Anginer (2008) finds the same result for the opposite side of the popularity specter, suggesting that companies that investors feel *"affectionate"* about, or admire, due to socially responsible policies or producing popular products is overbought, and thereby driving down future returns.

This is partially backed by Pastor et al. (2020) which uses an extension of CAPM by introducing *"ESG factors"* and *"climate betas",* which respectively incorporates unexpected changes in ESG concerns and firms exposure to climate shocks. "Green assets" are further defined as having satisfactory values of ESG-metrics, as opposed to "Brown assets". It is further shown that Green assets carry negative alphas due to the investors' individual non-financial preferences for these assets, as well as their ability to hedge climate risk (Chen, Pelger and Zhu, 2019). It is however shown that Green asset might outperform Brown ones when investors tastes shift towards Green assets. This shift in tastes or preferences result in a shift of real investments from brown assets to green ones, leading to positive social impact. Moreover, the alphas of the ESG investments are suggested to be lowest when investors preferences are at their most dispersed.

Moreover, NBIM (2021b) similarly distinguish between Green Assets and Brown Assets and investigate the effects of ESG-investing on two asset pricing models and come to similar conclusions as Pastor et al. (2020). The first asset pricing model is based on *"non-financial"* or *"ethical"* investing. That is, similarly to Pastor et al. (2020), the 'tastes' of individual investors are added into their model. This points to evidence that investors incorporate various assets into their portfolios of non-pecuniary motives. An ESG-efficient frontier, analogous to the efficient frontier schematically described in *Figure 2-1*, is then constructed, showing a market equilibrium with lower returns to Green assets than for Brown ones. The Brown assets increasingly outperforms Green ones when the fraction of ESG-investors in society increases relative to investors not incorporating ESG-factors into the investment decision. On the other hand, the model points to Green assets outperforming Brown assets in the short term as capital is flowing away from Brown assets and into Green ones, driving up prices and hence short-term returns.

The second proposed model considers *"risk-based ESG-investing".* That is, climate change related risks to assets' cash flows based on various future scenarios, including negative externalities on climate change from economic growth. Moreover, a fundamental principal in asset pricing found in university textbooks such as Cochrane (2001), claim that the utility of payoffs in bad economic times is higher than the same payoff in good economic times. This logic is applied by NBIM (2021b) to argue that since climate disasters are likely to cause bad economic times assets which pays off in climate disasters are viewed as less risky compared to assets which payoff only when such disasters are avoided. It follows from this model that Green assets might outperform Brown assets in severe climate scenarios and in the transition phase where the number of ESG-investors and the capital inflows in Green assets are rising but underperform in equilibrium.

Bolton *et al.* (2020) find that, controlled for size, book-to-market ratio, and momentum, companies with larger compared to lower CO_2 emissions and changes in CO_2 emissions earn higher returns. Additionally, institutional investors screen potential companies to invest in based on the intensity of direct CO_2 emissions. These results are further explained with investors' demand of compensation to carbon emission risk.

2.3 Empirical studies and machine learning in finance

In recent years, the volume of available data in the world of finance has increased tremendously. More data on various topics also allows for including larger amounts of data in economic models, both in volume and scopes. Simultaneously the availability of computational power and open-source software for simple implication of sophisticated models has also seen a phenomenal development. The more sophisticated machine learning models has shown superior predictive power compared to simpler models, particularly when examining more complex relationships and a larger number of factors. (Chen, Pelger and Zhu, 2019; Weigand, 2019)

Machine learning methods can be utilized to explore possible factors contributing to explanatory power. For example, Feng *et al.* (2020) utilize a combination of previously proven methods to find the marginal effect of each explanatory variable in a model on the expected returns of a cross-section. Furthermore, tree-based models are utilized by Bryzgalova, Pelger and Zhu (2019), and show that these are outperforming simpler models due to the ability to decode the complex relationships.

Overall, it is found that machine learning models can be utilized to cope with problems of high dimensionality and achieve better return predictions. The best performing models seem to be Neural Network models, as well as tree based models (Chen, Pelger and Zhu, 2019; Bryzgalova, Pelger and Zhu, 2020). The downside of these models is that they are hard to interpret, but this can be overcome by methods such as Shapley Additive exPlanations (SHAP) as done in Gradojevic and Kukolj (2022).

3 Data

The sample contains most of the companies listed in the S&P-500 over the period 2012-2021 and was compiled into a pooled longitudinal cross-section to be compatible with machine learning methods. The data was predominantly retrieved from Refinitiv Eikon, supplemented with data from Wharton Research Data Services (WRDS). Information on Federal Funds Effective Rates are retrieved from Federal Reserve Economic Research (FRED). The total sample size is 4906 observations. However, approximately 2% of the companies included in the S&P 500 are missing from the dataset, due to retrieval difficulties related to delisting's, mergers, and acquisitions. The sampling period is chosen mainly due to lack of available ESG-data prior to 2012. However, also within this period, the quality and frequency of especially ESG-data reporting is highly limited, and increasingly so further back in time, as illustrated in *Table 3-1*. The increasing number of missing observations back in time might lead to a *"selection bias"*, as companies that have been delisted or gone through mergers or acquisitions are underrepresented in the dataset. This is most prominent in the first couple of years where approximately 5% of the companies are missing.

Variables	Number of Missing Observations Per Year										
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Total
Absolute change Federal Funds Rate	0	0	0	0	0	0	0	0	0	0	0
Level Federal Funds Rate	0	0	0	0	0	0	0	0	0	0	0
Relative Change Federal Funds Rate	0	0	0	0	0	0	0	0	0	0	0
Year	0	0	0	0	0	0	0	0	0	0	0
EBITDA Margin	2	5	4	21	3	3	2	1	1	0	42
Environment Management Team	12	12	10	8	8	5	5	4	6	4	74
Environment Management Training	12	12	10	8	8	5	5	4	6	4	74
Environmental Materials Sourcing	12	12	10	8	8	5	5	4	6	4	74
ESG Score	12	12	10	8	8	5	5	4	6	4	74
Environmental Innovation Score	13	12	10	8	8	5	5	4	6	4	75
Environmental Pillar Score	13	12	10	8	8	5	5	4	6	4	75
Community Score	13	12	10	11	8	5	5	4	6	4	78
Policy Emissions Score	13	12	10	19	8	5	5	4	6	4	86
Full-Time Employees	7	7	9	5	4	8	8	6	17	17	88
Enterprise Value	12	12	10	8	5	4	4	4	15	16	90
EBIT Margin	2	5	4	75	3	3	2	1	1	0	96
Policy Environmental Supply Chain	12	12	10	30	8	5	5	4	6	4	96
Enterprise Value/EBITDA	13	15	13	8	8	12	8	5	5	20	107
Policy Sustainable Packaging	12	12	10	105	8	5	5	4	6	4	171
Revenue	24	20	20	18	15	12	14	10	20	21	174
Enterprise Value-to-EBIT	17	19	18	8	25	28	15	6	7	32	175
Corporate Social Responsibility Strategy Score	12	12	10	118	8	5	5	4	6	4	184
Total Debt to Total Equity	14	16	15	8	20	20	22	26	26	23	190
Price-to-Cashflow	21	22	14	8	26	21	14	14	19	43	202
Capital Expenditures	21	23	21	20	20	22	21	19	31	31	229
Price-to-Book	25	25	22	8	23	23	22	28	29	24	229
Long Term Debt-to-Total Capital	25	26	23	30	24	24	24	24	23	24	247
Product Responsibility Score	13	12	10	189	8	5	5	4	6	4	256
Income Available to Common Shareholders	25	26	24	136	25	25	25	24	24	24	358
Targets Emissions Score	17	19	16	260	13	10	11	9	11	12	378
Effective Tax Rate	28	30	24	8	45	58	43	34	39	70	379
Resource Reduction Policy	13	12	10	316	8	5	5	4	6	4	383
Resource Use Score	13	12	10	317	8	5	5	4	6	4	384
Price-to-Earnings	38	41	27	8	55	61	38	33	39	64	404
Policy Energy Efficiency	12	12	10	349	8	5	5	4	6	4	415
Targets Energy Efficiency	15	15	11	342	8	5	6	4	6	4	416
Price-to-Operating Cash Flow	17	24	18	260	20	15	17	9	13	26	419
Price Momentum	79	73	70	57	40	30	25	18	24	20	436
Targets Water Efficiency	12	12	10	406	8	5	5	4	6	4	472
Return on Total Assets	2	5	6	448	3	4	2	3	1	0	474
Resource Reduction Targets	13	12	10	460	8	5	5	4	6	4	527
Policy Water Efficiency	12	12	10	499	8	5	5	4	6	4	565
Toxic Chemicals Reduction	12	12	10	501	8	5	5	4	6	4	567
Equity Risk Premium	36	30	25	20	19	15	11	9	6	480	651
Price-to-Free Operating Cash Flow	61	81	86	14	81	80	73	55	67	70	668
Price-to-Common Equity	25	26	22	448	23	23	22	28	29	24	670
Dividend Yield Common Stock Primary	12	34	47	174	36	73	92	90	93	116	767
Tot Debt Cap-to-EBITDA	43	43	44	443	37	44	43	42	35	54	828
Historic Net Debt-to-Enterprise Value	129	134	125	8	90	85	82	81	86	91	911
Historic Price/Dividends	110	110	96	8	94	95	97	91	94	117	912
Price-to-Free Cash Flows	101	102	105	103	105	104	103	97	92	88	1000
Dividend	145	148	148	2	124	118	120	110	108	112	1135
Gross Profit Margin	108	101	118	254	117	110	100	121	115	114	1258
SDG 7 Affordable and Clean Energy	321	298	247	253	118	58	30	5	6	4	1340
CO2 Equivalent Emissions Total	163	190	196	14	186	155	138	103	99	111	1355
Total CO2 Equivalent Emissions to Revenues USD in million	164	191	198	30	191	155	138	103	99	111	1380
Beta	501	501	501	23	19	15	11	9	6	2	1588
CO2 Equivalent Emissions Direct, Scope 1	187	234	244	23	218	187	156	118	112	120	1599
Tax Rate - Actual	501	449	50	402	44	62	52	40	28	54	1682
CO2 Equivalent Emissions Indirect, Scope 2	220	260	252	2	227	193	166	126	115	124	1685
Energy Use Total	286	279	271	8	239	220	201	173	167	170	2014
Total Energy Use To Revenues USD in million	288	280	273	8	242	221	202	175	169	173	2031
Water Use To Revenues USD in million	297	288	267	360	255	240	231	206	196	200	2540
CO2 Equivalent Emissions Indirect, Scope 3	259	318	339	38	312	290	281	240	235	243	2555
CO2 Equivalent Emissions Indirect, Scope 3 To Revenues USD in million	261	319	340	50	312	292	284	244	237	246	2585
VOC or Particulate Matter Emissions Reduction Score	252	252	255	230	265	271	271	273	274	281	2624
Waste Recycled To Total Waste	374	363	345	370	328	314	308	289	284	288	3263
Greenhouse gas Emissions Method	440	423	409	369	349	329	301	298	287	269	3474
Downstream scope 3 emissions Processing of Sold Products	437	431	407	15	398	398	398	398	340	332	3554
Downstream scope 3 emissions Investments	437	420	407	21	398	398	398	398	340	340	3557
Downstream scope 3 emissions Transportation and Distribution	437	431	407	21	398	398	398	398	354	340	3582
Downstream scope 3 emissions End-of-life Treatment of Sold Products	437	410	407	26	398	398	398	398	372	369	3613
NOx Emissions	410	408	406	60	404	394	399	398	393	398	3670
Downstream scope 3 emissions Use of Sold Products	437	431	431	96	398	398	398	398	354	354	3695
Total Renewable Energy	499	492	486	501	440	340	299	265	251	250	3823
Renewable Energy Use Ratio	466	458	455	229	435	404	388	365	351	350	3901
Renewable Energy Use Ratio Score	466	458	455	238	435	404	388	365	351	350	3910
SOx Emissions	412	412	410	480	408	400	403	403	401	408	4137
Upstream scope 3 emissions Purchased goods and services	430	420	420	420	420	420	420	420	420	420	4210
Upstream scope 3 emissions Capital goods	430	421	421	421	421	421	421	421	421	421	4219
Upstream scope 3 emissions Fuel- and Energy-related Activities	430	430	430	430	430	430	430	430	430	430	4300
Upstream scope 3 emissions Business Travel	430	440	440	440	440	440	440	440	440	440	4390
Carbon Intensity per Energy Produced	501	501	501	13	501	501	501	501	500	500	4520
Return on Net Operating Assets	501	501	499	191	499	499	501	501	501	501	4694
Upstream scope 3 emissions Waste Generated in Operations	430	501	501	400	501	501	501	501	501	501	4838
Upstream scope 3 emissions Transportation and Distribution	430	501	501	410	501	501	501	501	501	501	4848

Table 3-1. List of variables and associated missing observations per year from 2012-2021 and total across all years. Showing a larger number of missing observations in earlier years, where ESG variables generally have large numbers of missing observations, several with over half the observations missing. Variables marked with grey are the ones

selected for the second-generation (G2) variable set, where the included variable with most missing values is *"Water Use To Revenues USD in Million*" with 2540 missing observations. The notation *"-to-"* indicate ratios with the left-hand side as numerator and right-hand side as denominator. EBIT is short for Earnings Before Interests and Taxes, while EBITDA is short for Earnings Before Interest, Taxes, Depreciation and Amortization.

Furthermore, the market related information such as price and market capitalization utilized to create suitable ratios are reported monthly. On the other hand, characteristics found in either the balance sheet, income statement or cash flow statements originates from the quarterly reports. Finally, most of the ESG-related data are reported annually. Therefore, the final dataset was measured at an annual frequency, as aggregating higher frequency data result in higher data quality than interpolating lower frequency data (Campbell and Thompson, 2008).

Initially, the data consists of 86 variables as given in *Table 3-1*. This initial set of variables is referred to as the first-generation variable set, using the abbreviation G1. It was selected based on data availability and on merits of feature importance from previous studies such as Black and Scholes, (1974), Fama *et al.* (1992), Hou *et al.* (2006), Penman, Richardson and Tuna (2007), Campbell and Thompson (2008), Arin, Mamun and Purushohman, (2009), Martani and Khairurizka (2009), Soliman, 2011; Holthausen and Zmijewski (2012), Gu, Kelly and Xiu, (2018b), Nissim (2019), Chen, Pelger and Zhu (2019) and Feng *et al.* (2020). As in Chen, Pelger and Zhu (2019), several variables assumed to be similar were included in the first-generation variable selection as the marginal difference could still contain important information. Selection of environment related ESG variables was quite similar to Silgjerd (2021), utilizing the same database with limited ESG variables.

Among the variables in G1, 34 are firm specific measures of profitability, size, investments, tax levels, volatility, and momentum. The remaining 52 variables are ESG-related, predominantly within the *"Environmental"* category, containing both qualitative and quantitative measures. Qualitative measures include true or false statements regarding whether a climate action has been taken by the company. The quantitative measures include both measured and estimated quantities of resource use and emissions, as well as scores provided by various rating agencies, compiled by Refinitiv Eikon.

Variables	Unit	Average	Std. Dev	Min	Max
52wk Total Return	%	15,20	34,00	-93,44	449,17
Year	Calendar Year	5,55	2,87	1,00	10,00
Relative Change Federal Funds Rate	%	41,26	78,30	-94,19	128,57
Level Federal Funds Rate	%	0,65	0,75	0,08	2,27
Long Term Debt-to-Total Capital	None	43,13	35,28	0,00	736,36
EBIT Margin	%	-375,12	15142,52	-749107,25	136,97
Income Available to Common Shareholders	%	-1254,35	121499,39	-7694272,73	3065406,03
Return on Total Assets	%	6,57	7,16	-50,41	48,45
Enterprise Value-to-EBIT	None	26,97	126,10	-13,24	7494,78
Price-to-Book	None	7,46	34,51	0,09	1077,28
Price-to-Cashflow	None	43,91	367,01	0,65	15257,33
Price-to-Free Operating Cash Flow	None	35,48	124,16	0,81	5657,62
Price-to-Operating Cash Flow	None	25,49	425,29	0,81	28198,43
Dividend	%	2,21	1,18	0,01	8,77
Tax Rate - Actual	%	24,81	48,91	-517,21	2994,44
Beta	None	0,92	0,47	-1,57	4,72
Price-to-Free Cash Flows	None	47,96	113,74	0,72	1864,38
Gross Profit Margin	%	47,83	20,52	2,47	100,00
Equity Risk Premium	%	5,88	0,44	3,83	8,82
Enterprise Value	$	61873605508,04	130060054948,3	-42972312660,96	2964572912070,01
Full-Time Employees	Full Time Equivalent	49458,68	122994,75	2,00	2300000,00
Capital Expenditures	$	1458603100,54	3078272436,90	40040,00	40140000000,00
Price Momentum	1-100 Score	57,57	28,11	1,00	100,00
Product Responsibility Score	1-100 Score	54,17	26,16	1,63	99,76
CO2 Equivalent Emissions Direct, Scope 1	Tonnes CO_2 Equivalents	3744967,71	11366446,31	3,00	135000000,00
Total CO2 Equivalent Emissions to Revenues USD in million	Tonnes CO_2-Equivalents/Million USD	343,65	933,19	0,00	9685,13
Energy Use Total	Gigajoules	31065152,50	123070101,71	27,97	2857000000,00
Water Use To Revenues USD in million	Cubic meters/Million USD	56228,49	585912,66	0,00	21283900,79
Community Score	1-100 Score	77,70	18,60	2,69	99,93
Corporate Social Responsibility Strategy Score	1-100 Score	57,42	26,90	0,21	99,67

Table 3-2. Descriptive statistic of the second-generation variable selection (G2). Describing the units of measurement, the average value, standard deviation as well as minimum and maximum values for each of the variables in G2.

Furthermore, a second-generation variable set (G2) displayed in *Table 3-2*. was obtained through utilizing XGBoost and Lasso regression to identify the most important ones from G1 as will be explained further in *Section 4.4.3*. One can further recognize from the table that none of the Boolean variables is included in G2, which suggests that they are not deemed particularly useful for explaining the cross-section of stock returns. One can also notice that *"EBIT Margin"* and *"Income Available to Common Shareholders"* have strong negative values, -375,1% and -1254,4%, respectively. This is due to outliers as can be seen by the extreme minimum values. If omitting the two smallest values for *"EBIT Margin"* and the smallest for *"Income Available to Common Shareholders"* the average values become more reasonable at 18,9%, and 11,1%, respectively.

Moreover, for each of the generations, the dataset was further split into an unrestricted subset of variables (UR), which contains all variables included in the respective generations, and a restricted subset (R), where the ESG-related variables are excluded. UR and R for G1 and G2 (G1UR, G1R, G2UR and G2R) are listed in *Table A-6* and *Table A-6* in *Appendix A*, respectively.

4 Methodology

This section lays out the methodology to explore the research question at hand. Starting with the test setting, describing the sequential data splitting, and testing, followed by a description of the regression models. Further, this section describes how the models are evaluated, before briefly explaining the key elements of the data pre-processing.

4.1 Test setting

In economics, statistical modelling is broadly applied to data with the aim of testing causal hypothesis. An example of this is the simple regression model, where various explanatory variables are assumed to cause an effect on the response variable. Further, predictive modelling can be viewed as the process of applying a statistical model to data for the purpose of predicting new or future observations (Shmueli, 2010; Beck, Hofman and Rohrer, 2018).

Out-of-sample estimates tells us how well our model performs on data outside the training data. The first step in this method is dividing the observations into training data and testing data. In this setting, the model is fitted/trained only to the training data. Next, in order to evaluate the out-of-sample performance, the fitted/trained models are evaluated on the observations in the testing set. This prevents data leakage, which occurs when the model predicts observations it has been trained on.

Much of the data is not likely to be publicly available at the time it was reported, particularly data from the income statement of balance sheet. Therefore, it can be wise to use lagged data (Fama *et al.*, 1992). Basu (1983) used a lag of three months, while Fama *et al.* (1992) used six months, however confessing this to be quite conservative, as even the companies not complying with the 90 day time limit before year end to submit their 10k, still would be available in April. Hence, similar to Lewellen (2014), a lag of 4 months is introduced, assuming that accounting data should be available four months after end of fiscal year. The required time for ESG-data to be available is not found in previous literature, but it is assumed to be the same as accounting data. It is worth noticing that data that are available much sooner than the four-month lag assumed for data from income statement or balance sheet, such as changes in stock prices and interest rates, will not reflect the immediate relationship between the explanatory variable and returns. It will however reflect asset prices, based on information investors received some time ago. The length of the lag will depend upon time it takes for the information to reach the market.

To compare the collective explanatory power of ESG variables the data is divided into an unrestricted variable selection (UR), containing all variables and a restricted variable selection (R), where the ESG variables are excluded.

As the relationship between the explanatory variables and the returns can change over time, the data is split into subgroups where the relationship is measured in each sub sample. For the in-sample analysis, the rolling window approach of five-year periods is applied, as illustrated schematically in *Figure 4-1*.

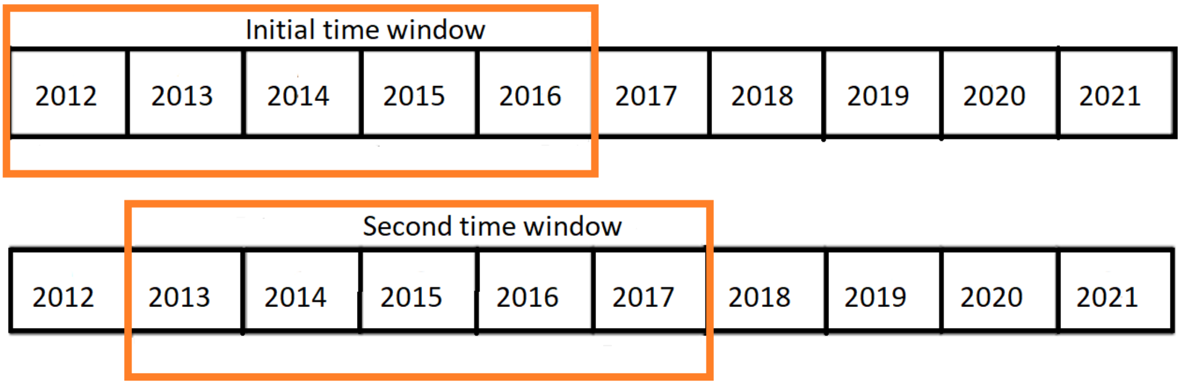

Figure 4-1. The rolling window approach where the data is divided into six distinct datasubsets consisting of six time windows. Each time window representing overlapping five-year periods. First time window (W1) from 2012-2016, W2 from 2013-2017, W3 from 2014-2018, W4 from 2015-2019, W5 from 2016-2020 and finally W6 from 2017-2021.

The first sub-sample containing data from 2012-2016, the second one from 2013-2017, third from 2014-2017 and so on, all the way up to 2021.

For the out-of-sample part of the analysis the data is split into a training set and test set, which is standard procedure in machine learning analysis. As the data is gathered from different points in time, it is important that the observations used for training the data is older than the ones used for testing, to avoid *"look-ahead bias"*. In this case, a four-year time window is utilized, where the first three years are training data and the last is testing data. In the first iteration data from year 2012-2014 was used as training data, while data from 2015 was used as testing data. In the next iteration, data from 2013-2015 was included in the training data, while the testing data was from 2016 and so on, as shown schematically in *Figure 4-2*.

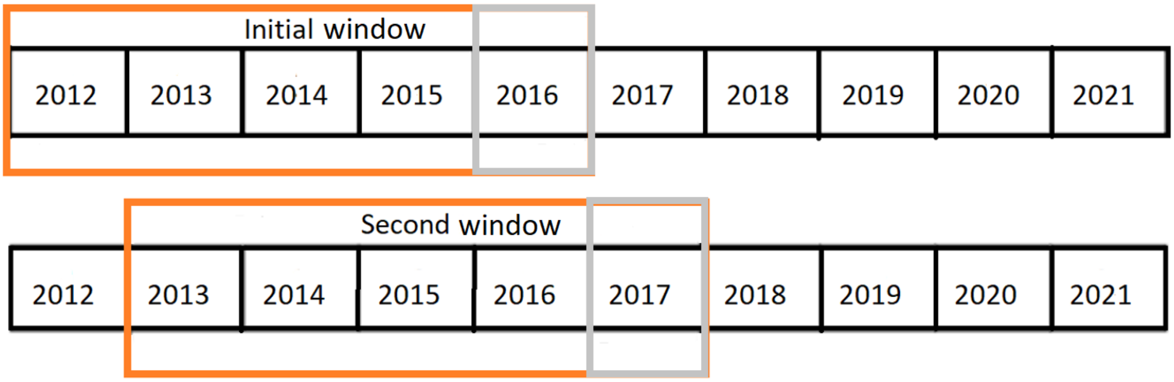

Figure 4-2. The rolling window approach where the data is divided into six distinct data subsets consisting of six time windows. Each time window representing overlapping five-year periods. Differing from *Figure 4-3* in that the first four years in each window is training data and the last one, marked with grey, is testing data.

4.2 Regression models

In this subsection the various regression models are explained. Starting the OLS method, its drawbacks and rationale for introducing popular machine learning models such as the Lasso and XGBoost regressions.

4.2.1 OLS

The OLS, as explained in detail in Wooldridge (2018a) is one of the most popular methods in empirical studies. It is a linear regression model which fits a line to the data by minimizing the sum of squared errors. It is suitable to ceteris paribus analysis as it allows directly controlling for many variables that simultaneously affect the dependent variable. These models are often referred to as multiple linear regression models (MLR), and one of the strong sides of MLR is the simplicity of interpreting the results. Since MLR is the only type of OLS-model used in this thesis, it will only be referred to as OLS from now on.

Let the general OLS model be expressed as

$$\hat{y}_i \ = \ \hat{\beta}_0 + \ \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \cdots + \ \hat{\beta}_p x_{ip} \tag{7}$$

where $\hat{y}_i$ is the estimated response for observation i, $x_1, x_2, \dots, x_p$ are the explanatory variables, and $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_p$ are the regression coefficients for each of the explanatory variables. The OLS model is trained by estimating the values of the coefficients $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_p$ such that the sum of squared residuals (RSS) over all n observations as given by the following equation is minimized:

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_j x_{ij} \right)^2 \tag{8}$$

where y_i is the actual observation of the response variable for observation i.

For the OLS estimators to be unbiased and consistent the unobserved error term must be independent of all explanatory variables for all companies at all times, mathematically expressed as:

$$E(u_i | x_{ij}) = 0$$

where u_i is the error term, being the difference between the estimated and actual response for observation i.

Whether this assumption is likely to hold in this thesis is very doubtful, by several reasons. One could for instance question the assumption that all explanatory variables affect the predicted returns in the same way. It could for example be argued that investors would think differently about CO_2 emissions coming from a large oil company opposed to a software company, as these operate under very different industry environments. By the same reasoning, it could also be assumed that the relationship between explanatory variables and the response variable would change over time. The latter will be controlled for by controlling for time-fixed effects as will be further explained in *Section 4.4*. Furthermore, the total set of Gauss-Markov assumptions required for the OLS estimators to be unbiased and consistent, as well as the best least squared estimator can be found in (Wooldridge, 2018a). The efficiency of OLS becomes highly unstable when incorporating a large group of parameters. Hence, it can be beneficial to narrow down the number of explanatory variables, for example trough utilizing XGBoost and Lasso regression.

4.2.2 The Lasso-regression

The Lasso (or Lasso-regression) is a model that is very similar to the MLR. Mathematically, the Lasso minimizes RSS, just as in the OLS-model, with an additional part that penalizes the number of variables in the model:

$$\text{RSS} + \lambda \sum_{j=1}^{p} |\beta_j| = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{9}$$

where the size of this penalty is proportional with the continuous hyperparameter λ often referred to as the penalty coefficient. Before estimating coefficients of the Lasso method, the input variables are standardized to have zero mean and a variance of one before performing the minimization in order to compare the variables on an equal scale.

Hence, the Lasso is a type of *"shrinkage"* method, where all the coefficients in the linear model is included initially and then *"shrinked"* towards zero based on their relative explanatory power as λ is increased. Ultimately, the estimated coefficients of variables with poor explanatory power are then lowered all the way to zero, and thus effectively excluded from the model (James, Witten, *et al.*, 2021). What is left is then a more parsimonious model, and the Lasso hence function as a natural variable selection method (Chatterjee, 2013).

In this thesis, for the Lasso-regression method, the tuning of the hyperparameter λ is determined through minimizing BIC by cross-validation of numbers between 0 and 10, with a step length of $0{,}01$. The resulting λ that minimized BIC differed some between the time windows. However, the average λ of $0{,}91$ was then chosen for all the windows.

Finally, where the OLS method is dependent on the first four Gauss-Markov assumptions to provide consistent and unbiased estimates, the Lasso requires fewer assumptions to provide consistent estimates, for example that the number of independent variables in fact can be higher than the number of observations. For more details on assumptions for the Lasso regression see (Chatterjee, 2013; Hlaváčková-Schindler, 2016).

4.2.3 XGBoost regression

The regression tree works in a very different way than linear models and have become very popular in machine learning. It is based on dividing the predictor space in many different segments based on an array of splitting rules. These splitting rules are based on the ones yielding the best results through minimizing the errors when the tree is trained. As these splitting rules can be schematically summarized in a tree-shape as demonstrated in *Figure 4-4*, these are often called decision trees, or regression trees when used for regressions as opposed to classification. The resulting partition space from the tree is also schematically shown in the figure. However, when utilizing more than two variables, the partition space will take multiple dimensions (Gu, Kelly and Xiu, 2020).

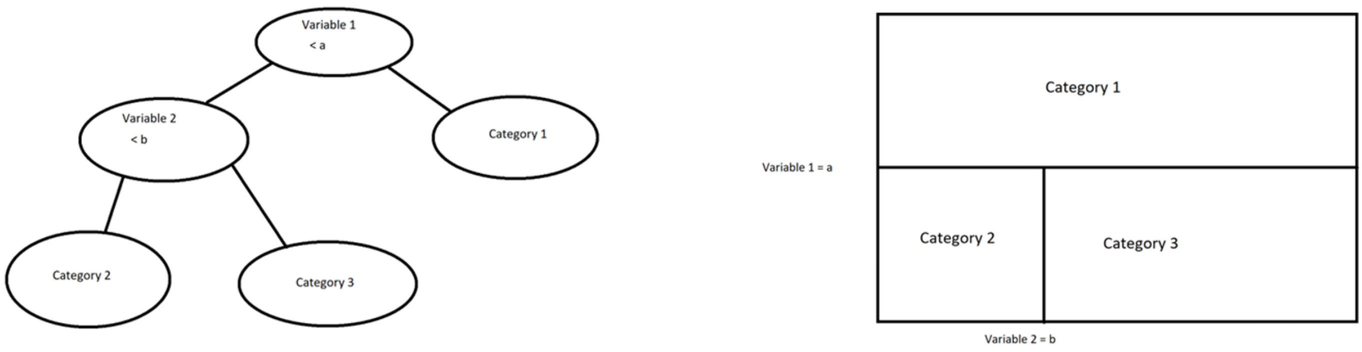

Figure 4-4. On the left-hand side of the figure a simple example of a decision tree is displayed. The tree is based on splitting criteria of two variables, a and b, leading to the accompanying partition space of three categories as showed on the right-hand side.

Furthermore, tree boosting is a widely used machine learning method that has proven to be highly effective. It incorporates the minimization of a regularized objective with a penalty term to help avoid overfitting, as in the Lasso regression described above. In the gradient boosted tree method, the trees are sequentially created based on the residuals from the previous one. Hence multiple decision trees are collectively used to generate trees with sequentially stronger predictive powers, ending up with one tree containing the *"consensus"* splitting rules based on the knowledge from the previous ones. The XGBoost method, presented by Chen and Guestrin (2016), is proven to be among the most effective and versitile tree boosting methods. The XGBoost method contains a variety of hyperparameters, and a good description of these can be found in (AWS, 2019) The hyperparameters was selected by cross validation using the *"GridSearchCV"* package imported from the Scikit-learn Python library (Sklearn, 2022), as recommended by Wang and Sun (2021). The initial values was based on the ones in Silgjerd (2021). The values of the initial and last iterations are displayed in *Table 4-1*.

Hyperparameter	Round 1	Round 6
Maximum Tree Depth	3, 4, 5	4
Learning Rate	0.01, 0.05, 0.5	0.075
Gamma	0, 0.25, 0.5	0.375
Regularization Lambda	0, 0.25, 0.5	0,0125
Minimum Child Weight	1, 8, 15	6

Table 4-1. Initial range and final hyperparameters used in the second generation XGBoost regression. Initial values recommended by Silgjerd (2021). Performing six iterations of cross validation. Sequentially updating ranges by keeping the selected value for each value and adding new minimum and maximum values, 50% smaller and larger than the middle one, respectively.

4.3 Model evaluation

4.3.1 The bias-variance tradeoff

To evaluate how good a model performs on the given dataset, an assessment of how well the predictions of the model fits the observed data in the dataset is in place (Felix, 2020). That is quantifying how close the predicted response values is to the true response value of given observations. Moreover, statistical methods can be manipulated to fit the data very well, by utilizing flexible methods and including a lot of explanatory variables in the model. However, increasing the number of explanatory variables come with the price of making the model more prone to overfitting the data. It would thereby not be representative of the true relationships, reflected by poor out-of-sample predictability.

Furthermore, as explained in detail in James, Hastie, *et al.* (2021) in statistical learning, there are two competing properties. It can be shown mathematically that the Mean Squared Error (MSE) for the test set can be decomposed into the squared variance, the squared bias and the variance of the error term, as follows:

$$E(y_0 - \widehat{f(x_0)})^2 = Var(\widehat{f(x_0)}) + [Bias(\widehat{f(x_0)})]^2 + \mathrm{Var}(\varepsilon) \tag{10}$$

The lefthand side of equation (10) represents the expected test MSE at a previously unseen test observation. This is the average MSE if the unseen observation was repeatedly tested utilizing numerous training data sets. The right-hand side of the equation suggests that in order to minimize the left-hand side, both relatively low variance and bias need to be obtained at the same time. Generally, increased flexibility, as measured by degrees of freedom, causes the variance to increase and the bias to decrease. When flexibility is increased from low levels, the decrease in the bias term tend to dominate the effect from the increase in the variance term, causing MSE to decrease. However, at higher levels of flexibility the increase in variance tend to dominate, causing MSE to increase. Therefore, the minimum test set MSE is found when the relative contributions from the two terms are equal.

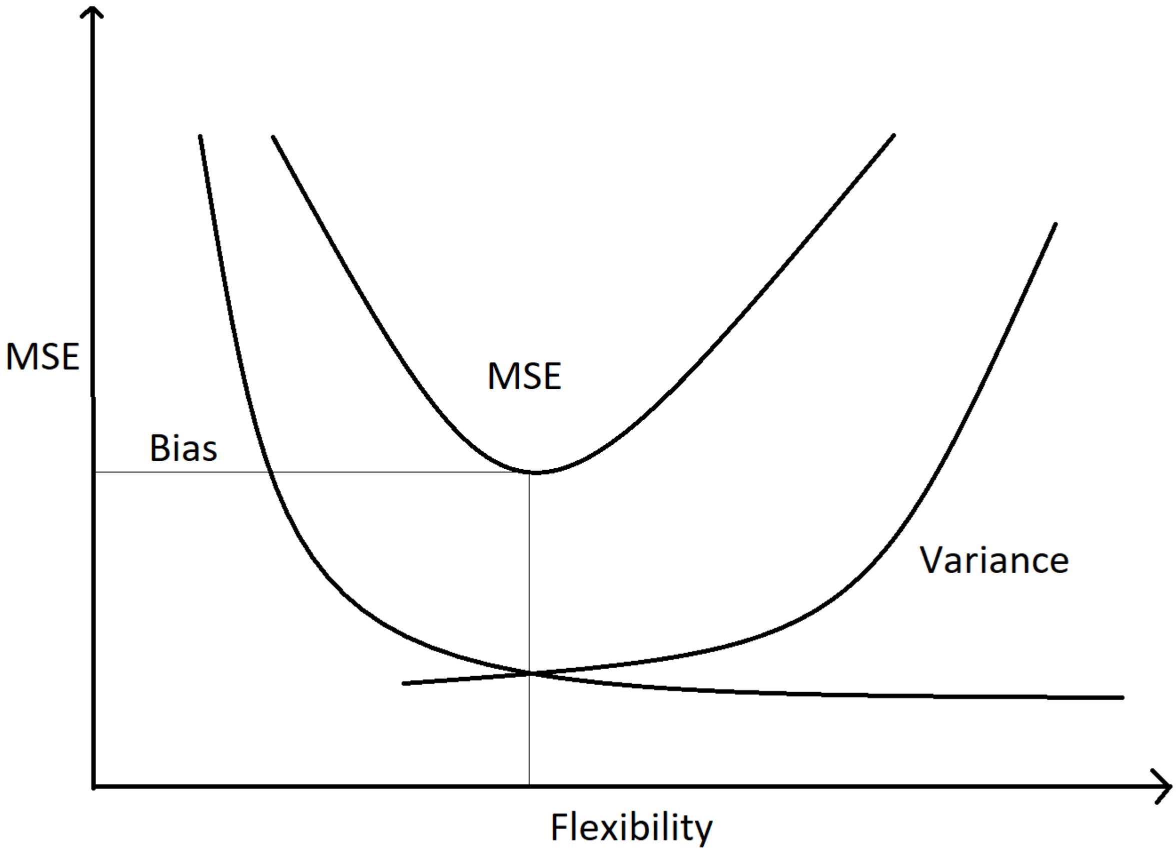

Figure 4-5. Illustration of the Bias-Variance Tradeoff, with the squared bias represented by the blur curve, the variance represented by the orange curve and the MSE represented by the red curve. Showing that the minimum MSE is found where the negative contribution from the MSE equals the positive contribution from the variance.

This tradeoff between low bias and low variance is referred to as the bias-variance tradeoff. The variance term refers to the degree in which the estimated response would change given that a different training dataset was used for estimation. High variance would occur if the estimation method were too flexible, and thus the estimated response would be highly sensitive to slight changes in the dataset. Bias on the other hand refers to estimating a relationship with a model not able to capture the true relationship between the explanatory variables and the response variable. For example, attempting to estimate non-linear relationships with a linear model (James, Hastie, *et al.*, 2021).

Followingly, it is useful to consider both how well the model fit the data, as well as the trade-off between reducing the sum of squared residuals and obtaining and overfitting the data. In this thesis, the former is measured by the Mean Squared Error (MSE) and R-squared (R^2). The latter is done by using a variety of techniques to adjust the in-sample error for the size of the model with respect to the number of explanatory variables included. These models can thus be utilized to select the *"best"* model in terms of having most desirable bias-variance tradeoff.

This will be done by implementing the Adjusted R^2, Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

In addition to examining how well the models perform, it can also be of great interest to investigate which of the variables that has the greatest impact on the model performance. That is, examining how each variable contribute to the model prediction and what is suggest for the causal relationship between the explanatory variables and the observed response values. This will be investigated trough regression coefficients for the OLS and Lasso Regression, as well as by Shapley Additive exPlanations (SHAP) values. Each of the metrics for examining in-sample goodness-of-fit, out-of-sample predictions, and variable importance will be defined and explained in the following sections.

4.3.2 MSE

The MSE is the RSS as derived in Equation (8) divided by the number of observations:

$$\text{MSE} = \frac{1}{n}\text{RSS} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{11}$$

So if the model does a good job of predicting the responses, then the difference between the actual observed responses and the predicted responses will be small, and so will the MSE (James, Hastie, *et al.*, 2021).

4.3.3 R^2

The total variability in the response variable about its mean is called the Sum of Squares Total (SST) and can be expressed mathematically as:

$$\text{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{12}$$

Where $\bar{y}$ is the mean of all observed responses. Further, the Sum of Squares Regression (SSR) is the variance in the response explained by the variance in the model, given by::

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \tag{13}$$

Finally, R^2 gives the percentage of variation in the response variable being explained by the model. It is given by (Campbell and Thompson, 2008) :

$$R^2 = 1 - \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{14}$$

4.3.4 Penalizing criteria

The Adjusted R^2 has a similar interpretation as the R^2. However, it penalizes additional explanatory variables in the model used to explain the same relationship. Hence including additional variables in the model with insufficient explanatory power will give a lower Adjusted R^2 Like the adjusted R^2, BIC and AIC introduce a penalty for additional explanatory variables in a similar fashion. Hence the best fit model according to these three criteria is the one explaining the largest fraction of the variance with the fewest explanatory variables. Both AIC and BIC is originally derived using log-likelihood, however when utilizing linear regression, the formula can be written as a function of the RSS, the number of explanatory variables (p) and the sample size (n), in the following equations (Narinç, Öksüz Narinç and Aygün, 2017):

$$Adjusted\ R^2 = 1 - \frac{(1 - R^2)(n - 1)}{(n - p - 1)} \tag{15}$$

$$BIC = n * ln\left(\frac{RSS}{n}\right) + p \ln(n) \tag{16}$$

$$AIC \ = n * ln\left(\frac{RSS}{n}\right) + 2p \tag{17}$$

More on assumptions, derivations and limitations of the BIC and AIC can be found in Kuha (2004), Ganatra, Panchal and Kosta (2010) and Hu (2012).

4.3.5 Regression coefficients

It is quite straight forward to interpret variable importance in regression models by interpreting the values of estimated coefficients: The magnitude of the estimated coefficient $\hat{\beta}_j$ of any variable in a linear regression model can be interpreted as the expected change in the response variable when increasing the variable by a one unit, holding all other variables constant. Further, the direction of this change is determined by the sign of the coefficient. How certain the resulting coefficients are, is determined by the *t*-scores with associated *p*-values and confidence intervals.

4.3.6 SHAP

Linear models are often preferred as they are easily interpreted, even though they are proven to perform worse than more complex models such as XGBoost, particularly with big data (Chen, Pelger and Zhu, 2019; Gu, Kelly and Xiu, 2020). In response, the framework of SHAP values was developed by Lundberg, Allen and Lee (2017), and has since gained popularity. The values explain the individual proportional impact from each variable to the prediction made by the model. They are in other words an interpretation of the difference between the base value that would be predicted if all features to the output of the current model were unknown, to the actual output made by the model. Further details on how Shapley values are estimated, including mathematical properties and derivations can be found in Lundberg, Allen and Lee (2017).

4.4 Preprocessing

4.4.1 Missing values

In the dataset gathered from Refinitiv Eikon, there were some missing data, particularly in the ESG-parameters that provided continuous data such as SOx emissions and *"Total CO_2 emissions to total revenues in USD dollars per tonne"*. There might be several reasons to why different businesses are lacking data on the various factors. An example is NOx- and Sox-emissions, which is the emissions of chemicals in the process of burning fuel, such as automobiles, and industrial sources such as power plants, industrial boilers and turbines (EPA, 2022b). One could assume that businesses that is not in the industrial segments or are large manufacturers or exporters would not be concerned with their NOx and SOx-emissions, and hence not spend effort in reporting them.

In this thesis missing values for continuous variables are imputation by k-Nearest Neighbors with $k = 4$. This method is based on replacing each of the missing observations with the $k = 4$ observations closest to it, measured by Euclidian distance in a predefined vector space. Further, missing values of dummy variables are one hot encoded to categorical values. Hence the single variable result in two different dummy variables, *"True"* and *"False"*, measured relative to the cases where the observation is missing. Similar utilization of KNN and one-hot encoding can be found in Sanjar *et al.* (2020)

4.4.2 Time-fixed effects

It is well documented that there exists a variety of macroeconomic variables that affect stock returns, where one of the most prominent one has shown to be interest rates, which is included by itself in the model (Eldomiaty *et al.*, 2020). However, one could also study the effect of other variables that vary across time but is common for all firms in a region. These can be Gross Domestic Product (GDP), inflation rates, commodity prices, conflicts, and uncountable other variables. As all companies included in the dataset are located in the US, one can to some degree assume that they are governed by the same macro variables, with some exceptions of local regulations in different states and cities within the US.

Hence, all other effects that are on the country level can be accounted for by a variable only varying across time and not companies, denoted by λ_t. Suggesting that the unobserved error term ν_{it} can actually be split into an idiosyncratic error term u_{it} that varies across both time and companies, and one time-fixed error term, denoted by λ_t. Moreover, introducing one dummy variable for each year can effectively control for λ_t, effectively eliminating it from the composite error term. This an example of the least square dummy variable model that can be described as:

$$y_{it} = \beta x_{it} + \lambda_1 D1_t + \lambda_2 D2_t + \lambda_3 D3_t + \cdots + \lambda_T DT_t + v_{it} \tag{18}$$

where D1 takes the value 1 for the first time period, and zero otherwise. D2 takes the value 1 in period 2 and zero otherwise and so on (Brooks, 2019a). This was not applied in XGBoost to avoid issues to obtain accurate variable importance scores as recommended by Kuhn and Johnson (2019), nor in the Lasso due to potential overfitting issues as recommended by Huang (2020).

4.4.3 Variable selection

Next, an XGBoost and Lasso regression was performed using the broad variable set, G1, as these methods handle high number of variables and multicollinearity. The output from the SHAP-plots and Lasso regressions for all six time windows are summarized in *Table A-2* and *A-3* of *Appendix A*, respectively. Only the variables among those with 20 highest SHAP values or selected by the Lasso regression for at least two time windows were included in the final variable set, marked by grey in the figures.

Moreover, the variables were further screened trough a multicollinearity analysis. Multicollinearity refers to one or several explanatory variables in the model being linearly dependent to one another (Wooldridge, 2018b). The presence of multicollinearity in the model does not break any of the standard Gauss-Markov assumptions, see Wooldridge (2018b) for details. However, even though it does not introduce bias or inconsistency in the model, multicollinearity makes it highly sensitive to slight changes of explanatory variables. Additionally, the significance and the sign of the regression coefficients become unreliable. The multicollinearity between the one-hot encoded variables from *"Year"* was ignored in this part of the variable selection process.

To mitigate the risk of multicollinearity the variables was also screened based the variance inflation factor (VIF) for all variables. VIF is calculated for each of the explanatory variables in the model, by performing a linear regression of the explanatory variable under investigation and each of the other variables in the model. The VIF is thus calculated as follows:

$$VIF = \frac{1}{1 - {R_p}^2} \tag{19}$$

where R_p^2 is the coefficient of determination between the variable under investigation and the other variables in the model. For a VIF score of 1 the variable shows no correlation with the other variables in the model. On the other hand, a VIF score of 2 would suggest that the standard error of the variable is $\sqrt{2} = 1{,}41$ times larger than if the variable was uncorrelated with all other variables in the model (Long *et al.*, 2018). This is the threshold recommended by Zuur, Ieno and Elphick (2010). The scores for all the second iteration variables are presented in *Table A-5*.

Furthermore, the correlation between each respective variable is displayed in *Figure A-1*. This was further utilized to identify the correlated variables causing the large VIF scores. A suitable cutoff VIF score of 3 inspired by Zuur, Ieno and Elphick (2010) was chosen. The four pairs of correlating variables causing the high VIF scores are presented in *Table A-4*. The variables of each pair can be assumed to contribute to explaining the cross-section of returns in quite a similar manner. Moreover, the presence of both variables in each pair is interfering with the value of the coefficients as well as its associated *p*-value. Hence, only one variable of each pair was selected in the final model. Which of the variables in each pair that was finally selected was determined through running an OLS regression. Here the variables in the correlation pair under investigation was excluded one by one and the variable with the lowest p-value in most windows was selected. This final variable selection is presented in *Table A-7*.

5 Results and Discussion

Table 5-1 presents five regressions that are conducted in this study, with their respective variable sets used for each method. In *Section 5.1* the results from these regressions will be presented, analyzed, and compared with respect to their relative qualities, measured by R^2, MSE, adjusted R^2, BIC and AIC. First assessing the relative goodness-of-fit scores of the variable sets G1 and G2 obtained from utilizing XGboost regression. Second, comparing the relative performance of UR and R for the three regression methods. Third, an in-depth analysis of ambiguous results from the penalizing criteria will be performed using basic calculus. Fourth, comparing the relative goodness-of-fit scores trough time. Finally, investigating the tradeoff between high R^2 and overfitting using the out-of-sample R^2.

Method	***Subsets performed on***
Lasso in-sample	*G1UR, G1R*
XGBoost in-sample	*G1UR, G1R, G2UR, G2R*
OLS in-sample	*G2UR, G2R*
XGBoost out-of-sample	*G2UR, G2R*
OLS Out-of-sample	*G2UR, G2R*

Table 5-1. List of regressions with variable subsets used for each method

In *Section 5.2* the variables that are suggested to have the best explanatory power will be presented and interpreted. This through analyzing the consistency of resulting regression coefficients and SHAP-plots with expectations from existing theoretical and empirical literature.

5.1 Goodness-of-fit

5.1.1 XGBoost

Table 5-2 present the goodness-of-fit metrics R^2, Adjusted R^2, MSE, BIC and AIC for the XGBoost regression for G1 and G2, in the left- and right-hand side of the table, respectively. The field marked by Δ represents the difference between UR and R for all the six time windows W1-W6, where W1 is the first one from 2012-2016 and W6 is the last one from 2017-2021.

XGBoost : G1	2012-2016			2013-2017			*XGBoost:* G2	2012-2016			2013-2017		
	UR	R	Δ	UR	R	Δ		UR	R	Δ	UR	R	Δ
R^2	0,120	0,116	0,005	0,118	0,116	0,003	R^2	0,115	0,112	0,003	0,117	0,110	0,007
Adj R^2	0,082	0,102	-0,020	0,081	0,103	-0,022	Adj R^2	0,104	0,103	0,001	0,106	0,101	0,005
MSE	556,240	559,310	-3,070	544,620	546,200	-1,580	MSE	559,400	561,600	-2,200	545,400	549,900	-4,500
BIC	-18351,0	-18835,0	484,0	-18533,0	-19023,0	490,0	BIC	-18873,7	-18918,9	45,2	-19065,7	-19100,5	34,8
AIC	-18921,2	-19037,0	115,8	-19103,3	-19226,2	123,0	AIC	-19047,0	-19052,0	5,0	-19239,7	-19233,8	-5,9
	2014-2018			2015-2019				2014-2018			2015-2019		
	UR	R	Δ	UR	R	Δ		UR	R	Δ	UR	R	Δ
R^2	0,109	0,105	0,004	0,204	0,203	0,001	R^2	0,106	0,104	0,001	0,200	0,200	0,001
Adj R^2	0,072	0,092	-0,020	0,170	0,192	-0,022	Adj R^2	0,095	0,096	-0,001	0,191	0,192	-0,001
MSE	495,820	498,420	-2,600	518,100	518,800	-0,700	MSE	497,960	498,660	-0,700	520,400	521,100	-0,700
BIC	-18635,0	-19126,0	491,0	-19061,0	-19556,0	495,0	BIC	-19167,7	-19218,8	51,1	-19587,0	-19639,0	52,0
AIC	-19212,0	-19330,0	118,0	-19632,0	-19759,0	127,0	AIC	-19341,8	-19352,3	10,5	-19761,5	-19772,2	10,7
	2016-2020			2017-2021				2016-2020			2017-2021		
	UR	R	Δ	UR	R	Δ		UR	R	Δ	UR	R	Δ
R^2	0,423	0,409	0,014	0,448	0,432	0,016	R^2	0,410	0,413	-0,003	0,438	0,432	0,006
Adj R^2	0,399	0,400	-0,001	0,425	0,424	0,001	Adj R^2	0,403	0,408	-0,005	0,431	0,427	0,004
MSE	924,210	946,350	-22,140	918,150	944,060	-25,910	MSE	944,800	939,700	5,100	935,000	944,000	-9,000
BIC	-19968,0	-20415,0	447,0	-20171,0	-20616,0	445,0	BIC	-20458,6	-20526,5	67,9	-20679,5	-20710,0	30,5
AIC	-20548,0	-20619,0	71,0	-20760,0	-20820,0	60,0	AIC	-20854,0	-20660,0	-194,0	-20854,0	-20844,0	-10,0

Table 5-2. Results from XGBoost for the variable set G1 (left) and G2 (right). Showing the various criteria scores and difference (Δ) between the Unrestricted variable set (UR) which includes ESG variables, and the restricted variable set (R) where they are excluded.

Starting with G2, the largest R^2 for both UR and R originates from W6, with scores of 0,438 and 0,432, respectively. The interpretation of the former is that about 44% of the total variability in *"52 Week Total Returns"* is explained by the variables in G2UR. The second largest scores for R^2 and adjusted R^2 come from W5, where the scores are marginally smaller than in W6. The third largest scores, however, are only about half of the two largest and come from W4. The scores from the three remaining time windows are all approximately one fourth the size compared to the largest two.

Comparing the R^2 of the G1UR on the left-hand side of the table, with G2UR on the right, the difference appears relatively small when recalling the difference in number of explanatory variables, 99 vs 30, respectively. The minimum difference in R^2 of 0,04% points is found in W4, and the maximum difference of 1,5% is points found in W6. Furthermore, recalling that higher values for adjusted R^2 and lower scores for BIC and AIC are favored, the adjusted R^2, BIC and AIC all favor both the G2UR over G1UR and G2R over G1R. Hence, the variables in G2 appear to outperform the ones in G1, suggesting the utilization of XGBoost and Lasso regressions for variable selection to be effective.

By the same logic as in the previous paragraph, the relative R^2, adjusted R^2, MSE, BIC and AIC seem to favor G1R over G1UR and G2R over G2UR. Suggesting that the ESG variables contribute positively to explaining the cross-section, however, not sufficiently to be a justified tradeoff for the additional variables.

What is also evident from *Table 5-2*, is that the MSE is notably larger for W5 and W6 than for the rest of the time windows, as seen for the case of R^2. This is somewhat counterintuitive considering that the R^2 decreases with an increase in the differences between the observed and predicted values, while MSE increases as described in *Equation (14)* and *(15)*, respectively. Hence, the only explanation for both R^2 and MSE being larger is if SST increases sufficiently to make up for the increase in (SSR) as described by *(14).* This is confirmed in *Table B-1* in *Appendix B*, showing that the SST values of 2016-2020 and 2017-2021, are indeed more than double the size compared to the other time windows.

To summarize, the results in *Table 5-2* imply that G1 provides better goodness-of-fit than G2 based on higher R^2 and MSE. The same argument applies for UR versus R in each of the variable selections. However, the higher R^2 seem to be a result of overfitting in both cases, as the Adjusted R^2, BIC and AIC, which penalizes additional variables, all favoring G2 over G1, and as R over UR. It is also apparent that the last two years provide a better goodness-of-fit than the rest, but this can be argued to be a result of fewer missing observations in the later years, providing a higher number of reported data, likely to perform better than the approximations provided through imputation.

5.1.2 OLS

The OLS regression was only performed on the G2 variables, by reasons expressed in *Section 4.2.1*. Each of the metrics in *Table 5-3* for the OLS is interpreted in the same manner as for XGBoost in the previous section. Similarly, W5 and W6 provide, by far, the largest R^2 and MSE. It is also evident that the R^2 is higher for the OLS regression in W1 and W2 for the G1 XGBoost, and in W1, W2 and W5 for the G2 XGBoost regression. Intuitively XGBoost should be expected to fit better to the data and provide larger R^2 in-sample due to its higher flexibility. However, the variable *"Year"* is included as an array of binary variables in the OLS regression, while one continuous variable in the XGBoost regression. This would suggest that the time-fixed effects as explained in *Section 4.4.2* to a higher degree is accounted for in the OLS, explaining the larger in sample R^2. Moreover, Adjusted R^2, BIC and AIC all favor R over UR in all six time windows. Suggesting that the inclusion of the ESG variables does not sufficiently increase the explanatory power when taking the penalty for introducing additional variables into account.

OLS: G2	2012-2016			2013-2017		
	UR	R	Δ	UR	R	Δ
R^2	0,139	0,138	0,001	0,137	0,135	0,002
Adj R^2	0,128	0,128	-0,001	0,124	0,135	-0,011
MSE	543,880	545,230	-1,350	533,400	534,500	-1,100
BIC	-18918,0	-18967,0	49,0	-19089,0	-19138,0	49,0
AIC	-19109,0	-19117,0	8,0	-19286,0	-19295,0	9,0
	2014-2018			2015-2019		
	UR	R	Δ	UR	R	Δ
R^2	0,105	0,102	0,003	0,188	0,184	0,004
Adj R^2	0,092	0,092	0,000	0,177	0,175	0,002
MSE	498,570	499,800	-1,230	528,580	531,520	-2,940
BIC	-19133,4	-19182,0	48,6	-19517,0	-19558,0	41,0
AIC	-19331,0	-19338,7	7,7	-19715,0	-19715,4	0,3
	2016-2020			2017-2021		
	UR	R	Δ	UR	R	Δ
R^2	0,417	0,415	0,002	0,427	0,426	0,001
Adj R^2	0,409	0,409	0,000	0,419	0,420	0,000
MSE	934,300	936,800	-2,500	959,190	954,500	4,690
BIC	-20455,0	-20503,0	48,0	-20503,0	-20652,0	149,0
AIC	-20652,7	-20660,0	7,3	-20801,0	-20809,0	8,0

Table 5-3. OLS for final selection of explanatory variables. Showing the various criteria scores and difference (Δ) between the Unrestricted variable set (UR) which includes ESG variables, and the restricted variable set (R) where they are excluded.

5.1.3 Lasso

Table 5-4 presents the goodness-of-fit results from the Lasso regression in similar fashion as for the previous two methods. However, even though the Lasso regression is performed on G1, the explanatory variables ending up with non-zero coefficients and effectively used to explain the cross-section is determined by the penalization process. Hence the number of explanatory variables (p) varies across time windows both for UR and R. The table further shows that the number of explanatory variables is effectively smaller for the Lasso than for the second iteration for OLS and XGBoost, between 14-21 for the UR and 8-18 in R, compared to 30 and 23 for the G2 XGBoost, and 38 and 31 for OLS.

Lasso: G1	2012-2016			2013-2017		
	UR	R	Δ	UR	R	Δ
R^2	0,094	0,083	0,011	0,081	0,070	0,011
Adj R^2	0,095	0,083	0,012	0,081	0,070	0,011
MSE	572,800	580,000	-7,200	568,000	575,000	-7,000
BIC	-18910,0	-19004,0	94,0	-19092,0	-19109,0	17,0
AIC	-19014,0	-19002,0	-12,0	-19173,0	-19156,0	-17,0
p	18	8	10	14	8	6
	2014-2018			2015-2019		
	UR	R	Δ	UR	R	Δ
R^2	0,041	0,032	0,009	0,149	0,145	0,005
Adj R^2	0,041	0,032	0,009	0,150	0,145	0,005
MSE	534,000	539,000	-5,000	554,000	557,000	-3,000
BIC	-19106,0	-19122,0	16,0	-19527,0	-19553,0	26,0
AIC	-19199,0	-19185,0	-14,0	-19632,0	-19626,0	-6,0
p	16	11	5	18	13	5
	2016-2020			2017-2021		
	UR	R	Δ	UR	R	Δ
R^2	0,386	0,385	0,001	0,299	0,293	0,006
Adj R^2	0,389	0,387	0,002	0,301	0,295	0,006
MSE	983,100	984,600	-1,500	1166,000	1175,000	-9,000
BIC	-20469,0	-20497,0	28,0	-20208,0	-20211,0	3,0
AIC	-20562,0	-20565,0	3,0	-20322,0	-20308,0	-14,0
p	16	12	4	20	17	3

Table 5-4. Showing the various criteria scores and difference (Δ) between the Unrestricted variable set (UR) which includes ESG variables, and the restricted variable set (R) where they are excluded. In *Table 5-2* and *5-3*, the number of variables used is the same across all time windows. In the Lasso on the other hand, the number of explanatory variables is reduced from 99 in UR to between 14-21 and from 36 in R to between 8-18.

Unlike the two former methods, the highest R^2 is found is W5, where it has been W6 for both OLS and XGBoost. However, the number of explanatory variables used for UR in W5 is 16, which is the second lowest for all the time windows. Hence, as expected, the value of adjusted R^2 is significantly higher, and BIC and AIC scores significantly lower than in the other windows. Suggesting that the 16 explanatory variables chosen by the Lasso Regression for W5 does a relatively good job of both explaining the response variable, while not overfitting it. When comparing the UR and R in W5, it is evident that UR is scoring marginally better than R in terms of R^2 and adjusted R^2 while R is favored by BIC and AIC in terms of lower scores. Hence, the adjusted R^2 is pulling in the opposite direction than BIC and AIC. Thus, which dataset proved the best overfit-underfit-tradeoff is inconclusive. For all other time windows than W5, the adjusted R^2 and AIC is favoring UR, while BIC is favoring R.

The Adjusted R^2, BIC and AIC for the G2 XGBoost, OLS and Lasso, is summarized in *Table 5-5*. Adjusted R^2, BIC and AIC favor the respective UR and R for the OLS over the others for

W1 and W2, and XGBoost for W3-W6. To summarize, the Lasso suggest the most parsimonious model, however it seems to overweight parsimony over explanatory power as measured by the penalizing criteria.

		Adjusted R^2		BIC		AIC	
		UR	R	UR	R	UR	R
W1	XGBoost: G2	0,104	0,103	-18874	-18919	-19047	-19052
	OLS	0,128	0,128	-18918	-18967	-19109	-19117
	Lasso	0,095	0,083	-18910	-19004	-19014	-19002
W2	XGBoost: G2	0,106	0,101	-19066	-19101	-19240	-19234
	OLS	0,124	0,135	-19089	-19138	-19286	-19295
	Lasso	0,081	0,070	-19092	-19109	-19173	-19156
W3	XGBoost: G2	0,095	0,096	-19168	-19219	-19342	-19352
	OLS	0,092	0,092	-19133	-19182	-19331	-19339
	Lasso	0,041	0,032	-19106	-19122	-19199	-19185
W4	XGBoost: G2	0,191	0,192	-19587	-19639	-19762	-19772
	OLS	0,177	0,175	-19517	-19558	-19715	-19715
	Lasso	0,150	0,145	-19527	-19553	-19632	-19626
W5	XGBoost: G2	0,403	0,408	-20459	-20527	-20854	-20660
	OLS	0,409	0,409	-20455	-20503	-20653	-20660
	Lasso	0,389	0,387	-20469	-20497	-20562	-20565
W6	XGBoost: G2	0,431	0,427	-20680	-20710	-20854	-20844
	OLS	0,419	0,420	-20503	-20652	-20801	-20809
	Lasso	0,301	0,295	-20208	-20211	-20322	-20308

Table 5-5. Summary table based on *Table 5-2, 5-3,* and *5-4*. Comparing adjusted R^2, BIC and AIC scores for the various methods across time windows.

5.1.4 Comparason of Favored Models

Moreover, it can be observed that the three selection criteria that includes penalization does in fact pull in opposite directions for the various regressions also when comparing UR and R. For the G2 XGBoost, the adjusted R^2 favors UR in W1, W2 and W6, R in the rest. AIC favors UR in W2, W5 and W6. The BIC on the other hand consequently favors R trough all time windows. Moreover, for the OLS regression adjusted R^2 favors UR in W4, R in the rest, while BIC and AIC favor R trough all time windows. While for the Lasso regression, the adjusted R^2 favors UR in all time windows, while AIC favors R in W5, UR in the rest. Contrarily, BIC favors R all time windows. Hence, it is clear that AIC, BIC and adjusted R^2 pulls in different directions in different situations, the mechanics of this phenomenon will now be examined.

	XGBoost: G1		XGBoost: G2		OLS: G2		Lasso: G1	
	UR	R	UR	R	UR	R	UR	R
Adjusted R^2	W6	W1-W5	W1,W2,W6	W3,W4,W5	W4	W1-W3,W5-W6	All	
AIC		All	W2, W5,W6	W1-W4		All	W1-W4, W6	W5
BIC		All		All		All		All

Table 5-6. Overview of in which time windows each of the criteria favors R or UR for the various regressions.

Since these criteria are solely based on mathematic formulas, the formulas themselves needs to be analyzed in order to explain the outcomes discussed above in the previous paragraph which is summarized in *Table 5-6*. When studying the formulas for adjusted R^2, BIC and AIC in *Equation (15)*, *(16)* and *(17)*, respectively, one can notice that the first part of the equations for BIC and AIC is identical. This part can be thought of as the one that favors low RSS, while the second part penalizes large number of explanatory variables trough *p*. Moreover, by partial differentiating the formulas with respect to *p*, one finds that:

$$\frac{\partial BIC}{\partial p} = \ln(\mathrm{n}) \tag{20}$$

$$\frac{\partial AIC}{\partial p} = 2 \tag{21}$$

$$\frac{\partial(Adjusted\ R^2)}{\partial p} = -\frac{(1-R^2)(N-1)}{(N-p-1)^2} \tag{22}$$

Hence, the increase in the respective scores from the marginal increase in the number of explanatory variables is $\ln(\mathrm{n})$ for BIC, to 2 for AIC and a more complex function of R^2, N and p for adjusted R^2. Meaning that the BIC imposes a larger penalty for additional explanatory variables than AIC for the data samples used in this thesis, where $\ln(\mathrm{n})$ is about $\ln(2500) = 7{,}82 > 2$ for the in-sample regressions. Testing the upper limit of *Equation (22)* by inserting the lowest observed value of R^2 and p, we get 0,0004. Hence, it is evident that the penalty in the adjusted R^2 is way lower than for both BIC and AIC in the datasets used in this thesis. This means that if more emphasis is put on parsimony and a smaller chance of overfitting, one should weigh the score of the BIC highest in the assessment. However, since the BIC is favoring R in all time windows, and one could then compare the values of the models, and the one with the lowest values of BIC would be favored. However, as Kuha, (2004) points out, one should try as far as possible to find models favored by both BIC and AIC.

5.1.5 Comparing scores trough time

As displayed by *Figure B-1, B-2* and *B-3* in *Appendix B,* the difference in the metrics scores between UR and R changes significantly over time. The change in the difference in R^2 between UR and R would suggest the relative contributions of ESG variables over time. If the ΔR^2 was strictly increasing over time, this would suggest that the dataset including the ESG variables was explaining an increasing fraction of the total variability, which would suggest that the impact of including the ESG variables was increasing over time. The opposite if the ΔR^2 was strictly decreasing over time. The results are however much more ambiguous than this. There does not exist any clear time trend for any of the metrics in the mentioned figures. However, the analysis is only based on six datapoints for each of the metrics, which makes the analysis highly sensitive to outliers. This is however a direct consequence of dividing the data into only six time windows. On the other hand, analyzing each year by itself would at the most extreme mean conducting a regression with 99 explanatory variables and only about 500 observations, which would require small-sample-adjusted equations for BIC and AIC. Hence another method for comparing the relative performance of the ESG variables over time is recommended.

5.1.6 Out-of-sample predictions

Table B-2 summarizes the out-of-sample R^2 scores for the OLS and G2 XGBoost regressions. What is immediately apparent is that all values of R^2 are in fact negative. This implies that the models perform worse than if the average value of all observations in the training set were used to estimate the returns. Hence the models are clearly overfit to the training data and are not suitable for explaining out-of-sample-relationships.

Opposed to the in-sample model, the R^2 of the model where ESG variable are excluded is higher for some windows than where they are included, in that they are less negative. This suggest that this model is less overfit, as would be expected by the same logic as it should result in a higher R^2 for in-sample explanations.

An explanation of the poor predictability as stated by Lewellen (2014) is that past cross-sectional slopes might be inaccurate estimates of the true future slopes due to noise in the estimates or due to time variation in the true parameters. These two mentioned effects may cause low out-of-sample predictive power, even though the firm characteristics historically would have been a significant predictor of returns. *Figure 5-1* illustrates the observed and predicted returns for the out-of-sample G2 XGBoost regression.

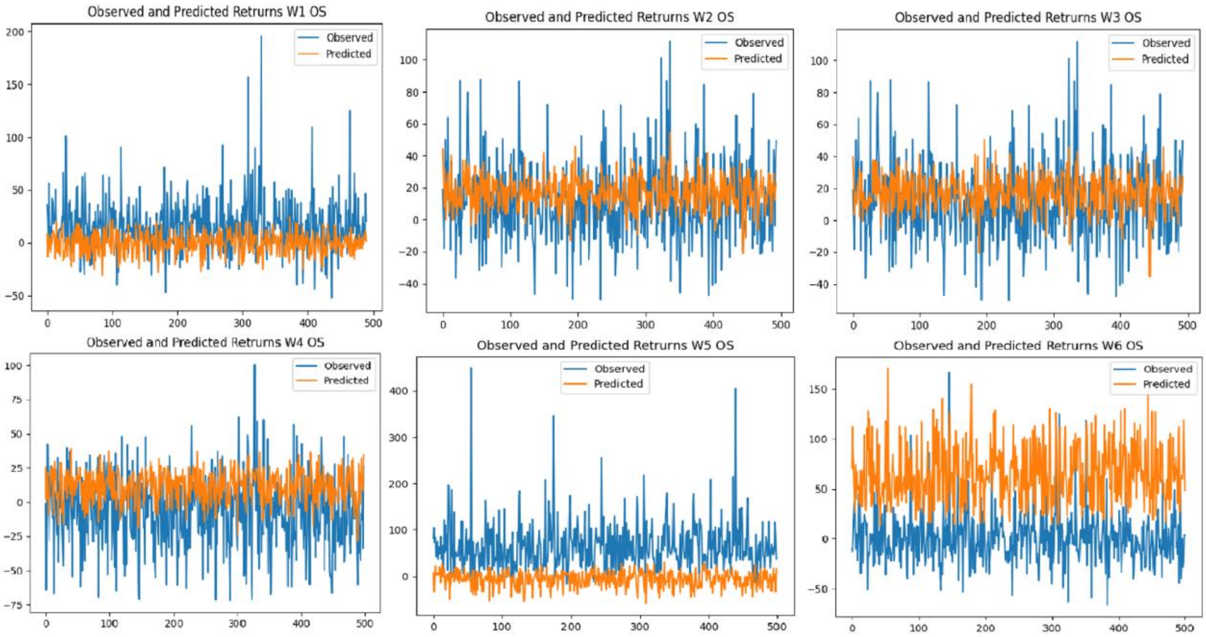

Figure 5-1. Observed and predicted returns for G2 XGBoost regression using the second-generation UR variable selection. Showing that the predictions indicated by orange seem to "shift" up an down relative to the observed values. Top Left W1, Bottom Left W2, Top-Middle W3, Bottom-Middle W4, Top-Right W5, Bottom-Right W6

There appears to be a *"Shift"* from the predicted values and the observed values. This can be a cause of using annual data, as the relationships between the explanatory variables vary across years but with too little variation in each time window for XGBoost to detect the actual relationship. Hence the implied relationships between the explanatory variables and returns from training the model on the first four years, might be very different from the fifth year containing the testing data. Moreover, the same information from the in-sample regression for the G2 XGBoost is illustrated in *Figure 5-2*.

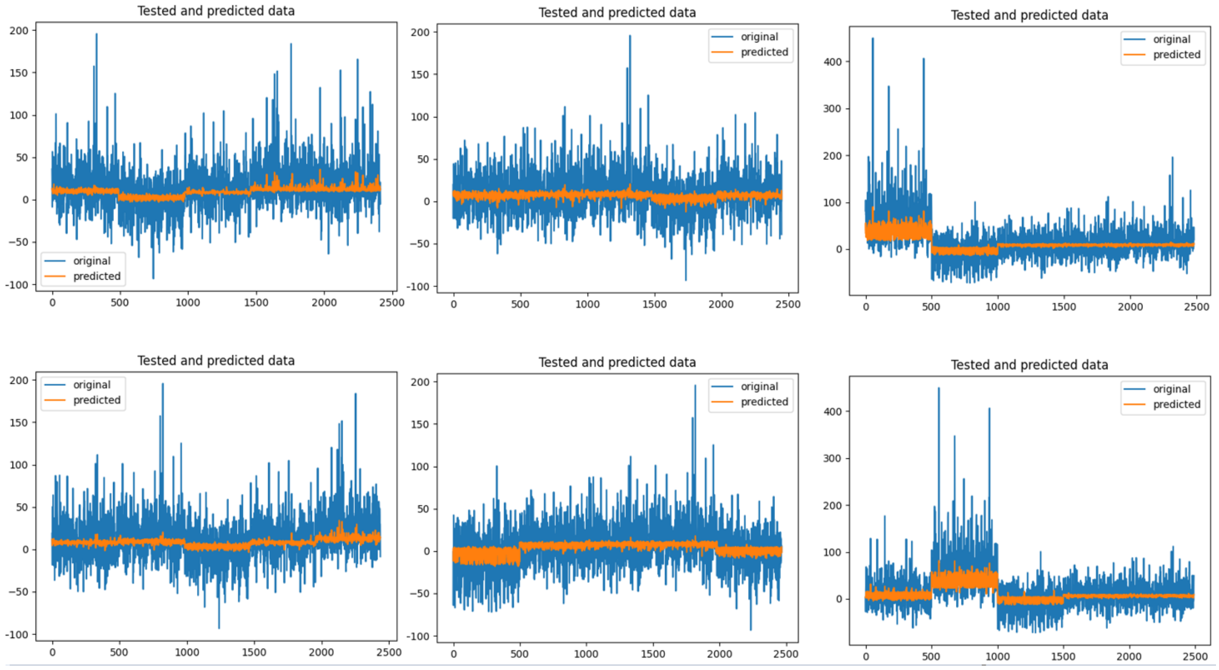

Figure 5-2. Showing the same as *Figure 5-1* but for the in-sample analysis. Now observing a clear shift for approximately each 500 observations, representing the 500 companies included for each year, indicating different relationships across time, which could explain the overfitting and poor out-of-sample scores. Top Left W1, Bottom Left W2, Top-Middle W3, Bottom-Middle W4, Top-Right W5, Bottom-Right W6.

From *Figure 5-2* one can clearly see shifts in the predicted values appearing for approximately every 500th observation. This suggests that the model overfits to the data specific to each year, such as the level and change in federal funds rates. The fact that these are equal for all companies in each year and vary across time mean that they effectively incorporate time-fixed effects in the same manner as the dummy variables for OLS as explained in *Section 4.4.2*. This would further mean that the effect of all unobserved variables that only vary over time but not across companies, such as GDP, unemployment rates, inflation rates and so on is mistakenly attributed to *"Relative Change Federal Funds Rate"* and *"Level Federal Funds Rate"*. This explains the overfitting on data for each year, as well as the counterintuitive results for several of the explanatory variables, which will be assessed in the final section of this thesis. Hence, one could argue that both *"Relative Change Federal Funds Rate"*, *"Level Federal Funds Rate"*, as well as any other time-fixed variables should be excluded when using an annual data frequency. This to avoid overfitting these variables by attributing all unobserved time-fixed effects to them. One could argue that these effects could be captured trough include a sufficient selection of time-fixed variables to the models. On the other hand, given the annual frequency of the data and the test setting in this thesis, only four and five of these would be used to fit the model, which is still likely to cause overfitting to each time window.

5.2 Variable importance

In this subsection, the SHAP-plot emerging from the G2 XGBoost regression and the coefficients with associated *p*-values from the OLS regression are used collectively to identify the most impactful variables in G2, interpret their relationship with returns and compare these with findings in existing theoretical and empirical literature.

Recall from *Section 4.4* that when using the OLS-regression, the different years are included as dummy variables and thereby incorporating time-fixed effects, effectively creating one indicator variable for each year. As listed in *Table C-1* in *Appendix C*, all the time-dummy variables are considered among the most prominent explanatory variables. Although, 2014 and 2021 are only appearing as significant in one window, they are still included as they are only present in a restricted number of the time windows in the first place. Additionally, *"Year"* was deemed highly important in both the Lasso and XGBoost regressions.

The generic interpretation of the OLS coefficient is that a one unit increase, where the units are given in *Table 3-2*, causes the expected response variable, *"52 Week Total Returns"* to increase by a number of percentage points given by the coefficient. The interpretation of the time-dummy coefficients are how the expected return changes given that the observations are obtained from the respective year, rather than the earliest one in each window. For instance, from *Table C-1* one can see that the coefficient of the variable *"2014"* for W1 is 1,78. Since *"52 Week Total Returns"* is given in percentage and that *"2012"* is the omitted dummy variable, the interpretation is that the expected total return is 1,78% higher, relative to 2012, given that an observation originates from 2014, all else equal. The significance levels show that there are significant effects from unobserved time-fixed effects. *Table 5-7* shows the actual annual returns of the equal weighted S&P 500 portfolio from 2012-2021. The table displays that for W1, relative to 2012, the expected total return increases by 18,14%, given that the observation originates from 2013, all else equal. However, according to *Table 5-7* one should assume that all dummies in W1 should be negative relative to 2012.

Calendar Year	Annual Returns for Equal Weighted S&P 500 Index
2012	17,9
2013	16,9
2014	11,1
2015	-2,2
2016	14,6
2017	9,3
2018	8,4
2019	-10,9
2020	54
2021	-0,6

Table 5-7. Actual annual returns, ex dividend, for the Equal Weighted S&P 500 Index from 2012-2021. Measured from April to April as in the dataset. Source: SPGlobal (2022).

Furthermore, one of the most important variables according to the SHAP-plot is *"Year".* This further supports that there are significant effects from unobserved time-fixed effects. When returning to *Figure 5-2*, one can observe that the shifts up and down is quite consistent with the actual returns presented in *Table 5-7.* The figure suggests that the predicted values for 2012 and 2013 are quite similar, with a small drop from 2013 to 2014, again from 2014 to 2015, and then a significant increase from 2015 to 2016. The general tendency of predictions shifting up in years of high returns, and down in years of low returns is also consistent for the rest of the time windows. The findings from *Figure 5-2* and *Table 5-7* are on the contrary not consistent with the signs of the dummy variables shown in *Table C-1*. For example, suggesting an 18,14% increase in expected returns in 2013 relative to 2012 all else equal, where the true value is lower, and a slight increase in expected returns of 1,78% for 2014 relative to 2012, when it the true value is about 7% lower.

This could suggest that the SHAP plot from the XGBoost model would be more reliable than the coefficients and p-values from OLS in determining variable importance and impact on predicted returns. This also supports the notion of problems of high dimensionality and multicollinearity, which is something the XGBoost model is less sensitive to.

Moreover, OLS suggests that *"Relative Change Federal Funds Rate"* is the most important variable. It is significant at a 5% significance level during all six time windows, with the lowest *p*-values in the last three windows, and the highest one in W2. An increasing rate should be expected to decrease the future value of a company and thereby making the future price of the company smaller than what it is today, therefore decreasing the returns in this time period. This through two mechanisms, the first one being that increasing rates are associated with a slower economy, decreasing the earnings and cash flows to firms. Additionally, the increased discount rate will decrease the present values of these cash flows (Jensen and Johnson, 1995; Bernanke and Kuttner, 2005). Moreover Guo (2004) finds that the market reacts negatively on unanticipated changes in the federal funds rate, but not to anticipated ones, as the anticipated changes would already be priced into the market. The same argument could be used to explain why *"Level Federal Funds Rate"* is less impactful. However, Blackrock (2021), further suggests that one should consider these traditional relationships between interest rates and stock prices in the light of the historically low levels of interest rates of the past decade. On the other hand, increasing, or high-interest rates are often a symptom of expected booms in the economy and associated inflation. If the interest rates would not be increased sufficiently to cool the economy down, this could increase prices and potentially also earnings of companies. Especially the ones able to allocate the increased cost of production over on consumers in the form of higher sales prices.

On the other hand, as explained in the previous paragraph, the *"Relative Change Federal Funds Rate"* and *"Level Federal Funds Rate"* are likely to incorporate unobserved time-fixed effects, therefore making the results unreliable. The same argument could be made for the equity risk premium, being a measurement of the market return over the risk-free rate, which is constant across companies but not across time, just as with *"Level Federal Funds Rate"* and *"Relative Change Federal Funds Rate"*. However, in theory, the equity risk premium should increase the excess return dependent on the beta of each company, as explained in *Equation (4)*.

Moreover, the effect of *"Beta"* as defined in the CAPM, should likewise depend on the equity risk premium. Yielding higher expected returns for higher values of *"Beta"*, given a positive equity risk premium. From the SHAP-plots of W1-W6 in *Figure C-1, C-2, C-3, C-4, C-5* and *C-6*, one can observe that the red and blue dots are spread in several clusters, whereas both colors appearing on both the left- and right-hand side, indicating negative and positive impacts on predictions, respectively. This insinuates that both high and low observed values can impact the predictions in both directions. Recalling that the Beta represents the sensitivity to shocks in the broad economy, one would expect a stock with high value to imply high expected returns in bull markets and low returns in bear markets. As seen both in *Table 5-7* and in *Figure 5-3*, the market has mainly provided positive returns during the past decade, but the realized stock returns are a noisy measure of expected stock returns (Elton, 1999).

However, Fama *et al* (1992) found that the Beta showed little explanatory power when controlled for size and book-to-market effects, the latter included inversely as *"Price-to-Book"* in this thesis.

Furthermore, *"Price-to-book"* is not deemed significant by the OLS, but the coefficients are negative and thus consistent with the findings of Fama *et al* (1992) in W3-W6. The rationale for the underperformance of companies with high *"Price-to-Book"* can be explained trough behavioral economics as done by Chan, Karceski and Lakonishok (2003). They argue that the cause may be that investors extrapolate the previous earning growth rates of firms, which leads to overvaluing companies that recently performed well. The pricing of these companies will then be reduced after a short period when the market corrects its mistake, causing companies with low book-to-market to underperform. It is also found in the same time windows in the SHAP-plots, where there is no clear structure in the clustering of red and blue dots, providing unclear results.

Moreover, as the betas reflect the sensitivities to changes in the broad economies, companies conducting more or less the same activities and are affected by the same type of market risk, tend to have similar betas. For example, companies in consumer goods tend to have low betas, while typical technology companies with large growth potential do tend to have high ones. Moreover, dependent on the macro environment, investors will prefer to hedge against certain risks. During time window W1 we know that oil prices plummeted in 2014, thus making the price and therefore returns on oil companies to fall a lot. Hence, a transitory preference for companies that happens to have a certain range of betas might explain why the higher levels of betas suggest lower returns in general bullish markets when one expects them to outperform companies of lower betas. Lastly, it should be noted that the values of the beta as estimated by Refinitiv Eikon should be assumed to be with some error, which would bias the coefficient downwards and the intercept upwards (Bodie, Kane and Marcus, 2020b). This problem should be assumed to apply to all variables that is not particularly accurately measured or estimated, especially variables such as CO_2 emissions and total water use.

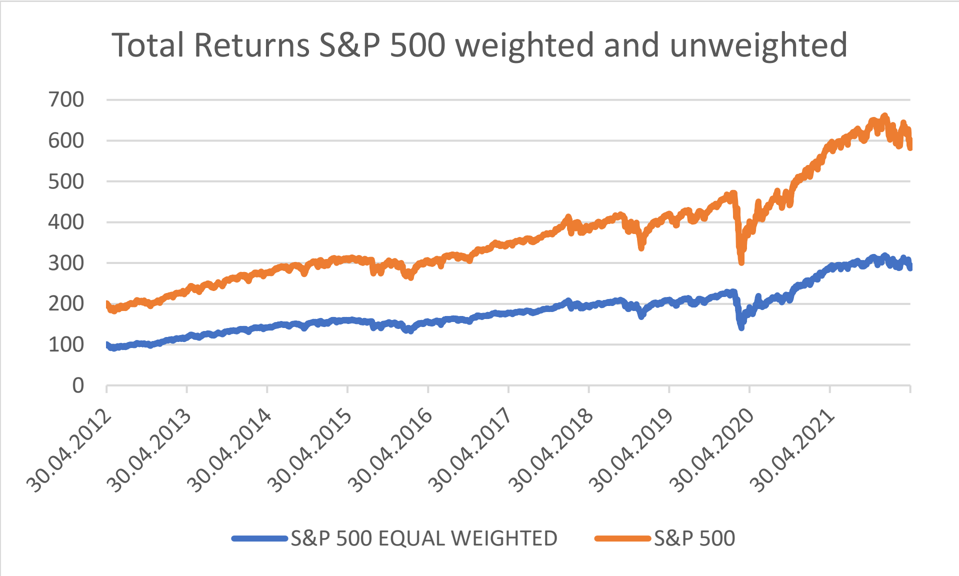

Figure 5-3 Total returns of the regular S&P 500 index, as well as the equal weighted index from April 2012-april 2022

Furthermore, the findings in this thesis are that there is a significant positive relationship between *"Enterprise Value"*, or *"size effects"* measured as market capitalization by Fama and French (1993), and returns. The coefficients are, however, very small, which is explained by the unit being *"Dollars"*, while a more convenient unit could be *"Millions of Dollars"* The current interpretation is that a one dollar increase in *"Enterprise Value"* leads to, for instance for W1, approximately to an increase of approximately 0,3 one-billionth percent increase in expected returns, all else equal. Hence to get more practical values, one could measure the effect of for example thousands, millions or even billions of dollars. One could also standardize all variables to have a standard deviation of one and zero mean, which is further explained in the final section. Nevertheless, the findings are inconsistent with the studies of Fama *et al.* (1992) and Banz, (1981) that both find there is a negative relationship between cross-sectional stock returns and the market cap of a company. The latter points to the fact that this effect is very large for very small companies, while the difference being milder when comparing average and very large companies. This could be part of the inconsistency, as all companies in the S&P 500 are at least average in size, and thus one should expect this effect to be milder. Furthermore, Merton (1987) suggest that the higher returns of smaller firms may be explained by large institutional investors neglecting these firms, and hence limit the quality of information available about these companies. This in turn increase the risk of the company being overpriced, and thus the investors require a higher risk premium for owning them. On the other hand Hou and Van Dijk (2019) refers to more recent studies from the 2000's pointing to a disappearing size effect, and even to a reversed effect, argued by negative profitability shock's to small firms.

Table C-1 further reports negative effects of the *"Price Momentum"* in four of the six time windows. This contradicts the findings in existing literature, such as Jegadeesh (1996), Fama and French (1996) and Hou *et al.* (2006), finds a significant positive relationship. As explained in *Section 3*, prices, which are updated daily is effectively lagged by 4 months relative to returns, which would put it in the window of 3-12 months were the momentum effects are shown to be especially significant. The premium for momentum is pointed to as being a result of irrational behavior driven pricing which could result in arbitrage opportunities. As pointed to by Fama and French (1993) companies that have recently experienced negative momentum, is also more likely to have low *"Price-to-Book"*, as well as having a smaller *"Enterprise Value"* as a consequence of this fall in stock prices. Hence it is further argued that the following abnormal reflects a higher risk premium and the factors functioning as proxies for some underlaying risk factors. Finally De Bond and Thaler (1990) argues that the anomalies previously discussed may also be an effect of human irrationality as explained in behavioral economics. Particularly that investors may overreact to information, as analysts may be incentivized to produce forecasts biased towards the more extreme to stimulate customer trades.

Furthermore, *"Price-To-Free-Cashflow"* is shown to have a positive relationship with returns. This is both counterintuitive and inconsistent with existing literature. See for example Hou *et al.* (2006), which finds a negative relationship, and argues that a high ratio would mean paying more for a free cash flow, which would be assumed to result in lower returns. Additionally, the G2 includes several of price and cashflows, including *"Price-to-Cashflow"*, *"Price-to-Operating Cash Flow"* and *"Price-to-Free Operating Cash Flow"*, which impacts would be interpreted in similar manner. Even though the selection method took multicollinearity into account, one could question the utility of including all of them in the variable set.

Moreover, the results for *"Capital Expenditures"* suggest negative relationships for all time windows, which are significant in four of them, which is consistent with existing literature such as Titman, Wei and Xie (2004). This is argued to be consistent with the notion that investors underreact to the notion of capital expenditures as *"empire building activities"*. Hence lower investments would yield larger free cash flows in the short term, and hence increase the ability to pay these cash flows out to investors as dividends in the short term. The coefficients are very small, which has the same explanation and solutions as discussed for *"Enterprise Value"*.

Moreover, the results from *"Tax Rate - Actual"* show a very small positive relationship for the significant coefficients, but it is only significant in W5 and W6. It also only appears in W5 and W6 in the SHAP-plot, but it is also here quite a strong positive relationship. When investigating whether a firms level of effective corporate tax rates affected a firms financial impact Brooks *et al.* (2016) found no discernible link between corporate taxes and stock returns. On the other hand, they found a negative impact of stock returns when a firm receiving negative media attention in the short term. Moreover, they did find that the trend of companies leaving a country in order to pay less tax is linked with more long-lasting falls in share price. This in agreement with the findings in this thesis, of being not highly impactful, but with a positive relationship between tax rates and stock returns. The fact that companies do not pay a lot of taxes could simply mean that it does not have a lot of taxable income, which would not be good. One could assume that for two companies with the same profits, it would be favorable to be able to pay less taxes, but as found in Brooks *et al.* (2016) being accused of tax avoidance is negative.

Moreover, one of the more significant results is the *"Dividend",* which is significant in all time windows for OLS, but negative. This is not consistent with, Shefrin and Statman (1984) suggesting investors prefer stocks with higher cash dividends. This arguably irrational preference might be due to behavioral economics, as investors might view dividends as money that can be spent, while not willing to sell shares from the same company. This despite the fact that most investors would have to pay more taxes for the cash dividends than on capital gain.

"Gross Profit Margin" is only significant in W1 and W4, where it is only positive in W4. This is both contra intuitive theoretically as one should assume that higher profits should lead to higher valuations of a company and thus higher returns. Profitability of a company has been empirically proven to have positive impacts on returns in a variety of studies, such as Basu, (1983), Fama *et al.*, (1992), Chan, Jegadeesh (1996) and E. Fama and French (1996). However, higher values of these metrics are only as good with respect to annual returns as their relative value in respect to what is expected by investors, as this is the base of the stocks pricing before new information is revealed. Recalling that Guo (2004) argues that it is predominantly the unexpected changes in the federal funds rates that affect stock returns, as the ones that are expected is already priced into the stocks. One should expect that the same would be true for essentially all risk factors, both firm specific and macro-factors. Instead, one could use rough estimates for expected values of earnings and other related parameters as

discussed in the previous paragraphs, trough measuring the average forecasts of various Wall Street analysts, and thus measure the *"earnings surprise",* as done by Rendleman, Jones and Latané (1982). As one would expect, they found that companies outperforming expectations experienced increasing prices and vice versa. What was unexpected however, was that the increases in price did not come immediately, as one should expect them to do. Instead, the positive-surprise stocks experienced abnormal returns after the information had become public. Which both suggests an anomaly in efficient market theory, as well as proposing that one should include lagged values of earnings-related variables to capture the entire cumulative effects.

None of the ESG variables are shown to be significant in the OLS model. However, *"Total CO_2 Equivalent Emissions-to-Revenues USD in Million"*, *"CO_2 Equivalent Emission Direct, Scope 1"*, *"Water Use-to-Revenues USD in Million"* and *"Energy Use Total"*, is included in several of the SHAP-plots. This might suggest that XGBoost recognize some complex nonlinear relationships between these particular ESG variables and returns, that are missed by OLS. The implications from the variables on expected total returns are, however, ambiguous in all cases, and marginally tilted towards suggesting that the more favorable ESG values have a negative impact on the predicted returns.

Energy use total is included in the SHAP-plots for W3, W4 and W6, showing a strong negative relationship with returns in the two former windows, and a weaker positive relationship in the latter. This would suggest that companies using less energy obtains higher total returns all else equal. However, this is very likely to be due to industry-specific measurements as well as strongly dependent on the size of the company and could be made more telling by both controlling for industries as well as size, for example by dividing by enterprise value or earnings as done for other variables. The same argument could be made of *"CO_2 Equivalent Emission Direct, Scope 1"*, where one would assume that the amount of CO_2 emitted should be heavily dependent on industry and size, as one should assume that oil companies would emit much more CO_2 than for example a software company with the same market value, and large companies more than small ones. The results for *"CO_2 Equivalent Emission Direct, Scope 1"* is however highly ambiguous, where there the dots seem more or less symmetrically clustered around zero. Moreover, *"Total CO_2 Equivalent Emissions-to-Revenues USD in Million"* indicate that high levels seem to have a negative effect on stock returns in W1 and W2, while being highly ambiguous in W4 and W6. The negative result would imply that companies responsible for less CO2 emissions per dollar earned would be expected to have higher returns, all else equal. However, one should control for the industry in which a company is, by the same logic as for *“Energy Use Total”*, and *“CO_2 Equivalent Emission Direct, Scope 1”* Finally, *“Water Use-to-Revenues USD in Million”* also provide ambiguous results, as high values seem to appear on the negative side, while blue dots are more frequently on the positive side, but also highly present on the negative side in W1-W4, while it seems to be quite strongly positive in W5. Once again, the argument of accounting for industry is also highly relevant in this case. One could interpret the vague clustering of the blue dots as there would be very much spread between the companies with relatively low water consumption, as would be expected to be in most industries, but more continuity in returns among companies with high water use. This could be as investors punish companies for using too much water relative to income, but it is once again, more likely to be a pure industry effect.

Hence the results from analyzing these four ESG variables would suggest that the companies with the better ESG scores have the lower returns. Further, recalling that the flow of capital from brown to green assets has been suggested to result in green assets outperforming brown ones in the short run. An effect that would be further enhanced by the momentum effect, as even investors not necessarily sharing the *“taste”* for the same companies but assume that the upward trend will continue. Leading the green companies to be overvalued and hence, increasing the prospects of carrying a negative alpha, suggesting future returns to be lower than the market return in equilibrium. Nevertheless, with increasing data availability also on ESG-data, combined with improved methods for modelling more complex datasets, identifying companies that have strong fundamentals in terms of climate risk ESG-factors could be done by analytics in similar fashions as done in fundamental analysis today. Having trusted analytics on the ESG-quality of a company could moreover provide more compelling evidence of a firm’s true ethical implications and climate risk mitigating properties than the current *“appearance based”* metrics. Leading the investors with ESG-preferences to allocate their capital to these companies, driving the prices, momentum and returns up in the near term.

6 Conclusion

In this thesis, a combination of machine learning regression models was applied on a cross-sectional dataset of firm characteristics, macroeconomic factors, and variables derived from the firms’ ESG reporting, to explain total returns on all individual stocks listed in the S&P 500 over the period 2012-2021. First, Lasso and XGBoost was utilized to select the most suitable variables from a variety of variables suggested by existing literature. These methods were seemingly effective as the selected variables outperformed the original set of all variables. Moreover, by utilizing XGBoost, Lasso, and OLS regressions on various data subsets, it became apparent that including the ESG variables provided some additional explanatory power. However, the advantage of including ESG variables is questionable when evaluating based on evaluation metrics that favor parsimony by penalizing additional variables, such as adjusted R^2, BIC and AIC. These criteria reached different conclusions for different models and for different time windows. Furthermore, the out-of-sample R^2 being negative for both the OLS and XGBoost implies that the relationships found by both methods are not likely to be representative of the true relationships, especially not ones that are constant over time.

The variables proved to have the highest impact on explaining the cross-section was to some degree consistent with previous literature. However, the change in federal funds rates, the beta, dividend yield, size, price-to-free-cashflows, tax rates, and momentum effects on returns showed ambiguous impacts on return predictions, or in the opposite direction of what should be expected by both theoretical and empirical literature. None of the ESG variables were shown to be significant in the OLS model. However, *“Total CO_2 Equivalent Emissions-to-Revenues USD in Million”*, *“CO_2 Equivalent Emission Direct, Scope 1”*, *“Water Use-to-Revenues USD in Million”* and *“Energy Use Total”* are included in several of the SHAP-plots. This might suggest that XGBoost recognize some complex nonlinear relationships between these ESG variables and returns, that are missed by OLS. The implications from the variables on expected total returns is, however, unclear in all cases, and marginally negative, suggesting that scoring well on ESG metrics have a negative impact on the predicted returns, all else equal.

Moreover, the reviewed literature confirms the finding of this thesis as it implies that variables based on ESG reporting does not have a particularly strong impact on stock valuations or investment decisions. The ambiguous, but slightly negative impacts of ESG variables on stock returns suggests that Green stocks is expected to provide lower rates of return than the market. Furthermore, ESG investing strategies might outperform the market in the short term due to increased capital inflow in Green stocks but underperform once the market reaches equilibrium.

Nevertheless, the empirical findings in this thesis are likely to be seriously compromised by several reasons. One of the most prominent ones is the large number of missing observations, particularly for the ESG variables. The problem of missing observations was attempted solved through KNN-imputation. Even so, the imputed values are likely to deviate from what would be the true values if the firms reported across all ESG measures. Further, some errors are likely to be present in the reported observations as well, especially the ESG variables, as these are less strictly audited. It is also likely that the model contains "omitted variable bias", both in terms of the idiosyncratic error term, but also a "heterogeneity bias" as the models does not control for the potential difference in relationships between the explanatory variables and returns for different companies. Moreover, the uneven distribution of missing data across time may cause a selection bias, as the most recent observations are overrepresented in the model compared to the least recent observations. Further, the test setting of utilizing six time windows with five years in each appear to not be a good setting when having annual data frequency. Particularly, as time-fixed variables such as the federal funds rate will incorporate the unobserved time-fixed effects, leading to overfitting and poor out-of-sample performance. Finally, even though the variables with the most significant contributions to multicollinearity was eliminated, it is likely that some of the remaining variables still suffer from some degree of multicollinearity. Even though Lasso and XGBoost handles multicollinearity well, this would still affect the estimated coefficients and p-values for the OLS. Fortunately, the data related issues, particularly concerning quality of ESG-data, might become less pressing in the future, as the quality is likely to improve due to increasing regulation and standardization of ESG-reporting. Moreover, the current issues experienced in this thesis could to some degree be mitigated by taking certain measures to better adapt the research methodology. Some specific proceedings will be recommended in the final section.

6.1 Recommendations for further work

As previously discussed, the main limitation of this thesis is data related issues, such as the quality of the ESG-data and the measurement frequency, leading to overfitting and a variety of biases. Attempting to obtain higher quality data, particularly with respect to the ESG variables, is recommended. This could be done by utilizing data from more databases. Furthermore, this would also increase the possibilities of obtaining a higher data frequency, for example by interpolating the annual data to monthly frequency. This is likely to help with the severe overfitting issues associated with time-fixed effects variables. Better data collection could also open for including an even broader initial selection of variables, including additional firm-characteristics, macroeconomic factors as well as other ESG variables. This should also include lagged and first differenced versions of certain variables, as the effects from certain variables are shown to be absorbed by the market over time. As investors are concerned not only with the level of certain variables, but also how they change over time, particularly if the change is surprising, differenced versions of certain variables should be included. The broader initial variable selection is likely to allow Lasso and XGBoost to identify additional relevant explanatory variables, as well as reducing the omitted variable bias associated with the idiosyncratic error term. Moreover, variables could be selected using the same methodology in this thesis, but with a stricter filter with respect to high dimensionality and VIF, helping to reduce the multicollinearity-related bias for OLS. However, one could utilize the method proposed in Feng *et al.* (2020), selecting factors that are useful either to explain the cross-section of expected returns or mitigating the problem of omitted variable bias du to selection mistakes. Obtaining reliable *p*-values could be further enhanced by utilizing cluster robust standard errors, handling both autocorrelation as well as heteroskedasticity in the error term.

Moreover, to ease comparison of the various variables representing a variety of factors, standardization should be utilized when performing the in-sample analysis. That is, transforming the dataset so that all variables have a mean of zero and a standard deviation of one. Further, to better assess how the collective impact of ESG variables for explaining returns changes over time, one could create reaction terms between the ESG variables and the year-dummy variables. Finally, controlling for different sensitivities to different industries is recommended to handle the heterogeneity bias. This could be done by introducing a separate dummy variable representing each sector. Alternatively, organize the data as panel, and utilizing the “fixed effects”-method as described in detail in Brooks (2019b) and Wooldridge (2018b).

List of References

Aluchna, M. and Roszkowska-Menkes, M. (2019) 'Non-financial Reporting. Conceptual Framework, Regulation and Practice', pp. 213–236. doi: 10.1007/978-3-030-00440-8_14.

Amran, A. and Ooi, S. K. (2014) 'Sustainability reporting: Meeting stakeholder demands', *Strategic Direction*, 30(7), pp. 38–41. doi: 10.1108/SD-03-2014-0035/FULL/XML.

Arin, K. P., Mamun, A. and Purushothman, N. (2009) 'The effects of tax policy on financial markets: G3 evidence', *Review of Financial Economics*, 18(1), pp. 33–46. doi: 10.1016/J.RFE.2008.05.001.

AWS (2019) *Amazon SageMaker - Developer Guide*. Available at: https://docs.aws.amazon.com/sagemaker/latest/dg/sagemaker-dg.pdf#xgboost_hyperparameters.

Banz, R. (1981) 'The relationship between return and market value of common stocks', *Journal of financial economics, 1981, Vol.9 (1), p.3-18*, 9.

Basu, S. (1983) 'The Relationship Between Earnings Yield, market value and return for NYSE Common Stocks', *Journal of Financial Economics*, 12, pp. 129–156.

Beck, E., Hofman, J. and Rohrer, J. (2018) 'Explanation, prediction, and causality: Three sides of the same coin?', pp. 1–14. Available at: https://osf.io/u6vz5/.

Bernanke, B. S. and Kuttner, K. N. (2005) 'What explains the stock market's reaction to federal reserve policy?', *Journal of Finance*, 60(3), pp. 1221–1257. doi: 10.1111/j.1540-6261.2005.00760.x.

Black, F. (1992) 'Beta and Return', *The Journal of Portfolio Management Fall 1993, 20 (1) 8-18*, pp. 74–77. doi: https://doi.org/10.3905/jpm.1993.409462.

Black, F. and Scholes, M. (1974) 'The effects of dividend yield and dividend policy on common stock prices and returns', *Journal of Financial Economics*, 1(1), pp. 1–22. doi: 10.1016/0304-405X(74)90006-3.

Blackrock (2021) *Why rising rates won't derail stocks*. Available at: https://www.blackrock.com/sg/en/insights/why-rising-rates-wont-derail-stocks (Accessed: 1 June 2022).

Bodie, Z., Kane, A. and Marcus, A. (2020a) *Investments: Chapter 11*. 12th edn. McGraw-Hill Education.

Bodie, Z., Kane, A. and Marcus, A. (2020b) 'Investments', *McGraw-Hill Education*, (12), p. 211.

Bolton, P. *et al.* (2020) 'Do Investors Care about Carbon Risk?', *Journal of financial economics, 2021-11, Vol.142 (2), p.517-549*. doi: 10.3386/W26968.

De Bond, W. F. . M. and Thaler, R. (1990) 'Do Security Analysts Overreact ?', *The American Economic Review*, 80(2), pp. 52–57.

Brooks, C. *et al.* (2016) 'Do investors care about corporate taxes?', *Journal of Corporate Finance*, 38(January 2000), pp. 218–248. doi: 10.1016/j.jcorpfin.2016.01.013.

Brooks, C. (2019a) 'Introductory Ecenometrics For Finance', *Introductory Ecenometrics For Finance*, (4), pp.

493–496.

Brooks, C. (2019b) ‘Introductory Econometrics for Finance - Fixed effects and time-fixed effects’, (4), pp. 490–496.

Bryzgalova, S., Pelger, M. and Zhu, J. (2020) ‘Forest through the Trees: Building Cross-Sections of Stock Returns *’. doi: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3493458.

Calhoun, M. A. (2017) ‘Profits before People: Perversion of the Friedman Doctrine’, *https://doi.org/10.5465/ambpp.2015.18560abstract*, 2015(1), p. 18560. doi: 10.5465/AMBPP.2015.18560ABSTRACT.

Campbell, C. and Thompson, S. B. (2008) ‘Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? Terms of Use Share Your Story’, *The Review of financial studies, 2008-07-01, Vol.21 (4), p.1509-1531*. doi: 10.1093/rfs/hhm055.

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

Chan, Jegadeesh, and L. (1996) ‘American Finance Association Momentum Strategies Author (s): Louis K . C . Chan , Narasimhan Jegadeesh and Josef Lakonishok Source : The Journal of Finance , Vol . 51 , No . 5 (Dec ., 1996), pp . 1681-1713 Published by : Wiley for the American Finance’, 51(5), pp. 1681–1713.

Chan, L. K. C., Karceski, J. and Lakonishok, J. (2003) ‘The Level and Persistence of Growth Rates’, *Journal of Finance*, 58(2), pp. 643–684. doi: 10.1111/1540-6261.00540.

Chatterjee, S. (2013) ‘Assumptionless consistency of the Lasso’, pp. 1–10. Available at: http://arxiv.org/abs/1303.5817.

Chen, L., Pelger, M. and Zhu, J. (2019) ‘Deep Learning in Asset Pricing’, *SSRN Electronic Journal*. doi: 10.2139/ssrn.3350138.

Chen, T. and Guestrin, C. (2016) ‘XGBoost: A Scalable Tree Boosting System’, *Proceedings of the 22nd ACM SIGKDD International Conference on knowledge discovery and data mining, 2016-08-13, Vol.13-17-, p.785-794*. doi: 10.1145/2939672.2939785.

Cochrane, J. H. (2001) *Asset Pricing Revised Edition*. Revised. Princeton University Press.

Divanbeigi, R. and Ramalho, R. (2015) ‘Business Regulations and Growth’, *World Bank Policy Research Working Paper No. 7299*. doi: 10.1596/1813-9450-7299.

Eldomiaty, T. *et al.* (2020) ‘The associations between stock prices, inflation rates, interest rates are still persistent: Empirical evidence from stock duration model’, *Journal of Economics, Finance and Administrative Science*, 25(49), pp. 149–161. doi: 10.1108/JEFAS-10-2018-0105.

Elton, E. J. (1999) ‘Expected return, realized return, and asset pricing tests’, *The Journal of finance (New York), 1999-08-01, Vol.54 (4), p.1199-1220*, LIV(4), pp. 257–278. doi: 10.1142/9789814335409_0013.

EPA (2021) *Regulations for Greenhouse Gas Emissions from Commercial Trucks & Buses | US EPA*. Available at: https://www.epa.gov/regulations-emissions-vehicles-and-engines/regulations-greenhouse-gas-emissions-commercial-trucks (Accessed: 25 April 2022).

EPA (2022a) *Deepwater Horizon – BP Gulf of Mexico Oil Spill*. Available at: https://www.epa.gov/enforcement/deepwater-horizon-bp-gulf-mexico-oil-spill (Accessed: 11 June 2022).

EPA (2022b) *Nitrogen Oxides Control Regulations | Ozone Control Strategies | Ground-level Ozone | New England | US EPA*. Available at: https://www3.epa.gov/region1/airquality/nox.html (Accessed: 3 April 2022).

EY (2021a) *Global IPO market has record-breaking 2021, prepare for headwinds in 2022 | EY - Global*. Available at: https://www.ey.com/en_gl/news/2021/12/global-ipo-market-has-record-breaking-2021-prepare-for-headwinds-in-2022 (Accessed: 25 April 2022).

EY (2021b) 'Mainstreaming ESG in the pre-IPO process', (December). Available at: https://r.search.yahoo.com/_ylt=AwrIdVfcnKlilTcAtFPsDYpQ;_ylu=Y29sbwNpcjIEcG9zAzEEdnRpZAMEc2VjA3Ny/RV=2/RE=1655311709/RO=10/RU=https%3A%2F%2Fassets.ey.com%2Fcontent%2Fdam%2Fey-sites%2Fey-com%2Fen_in%2Ftopics%2Fclimate-change%2F2022%2F01%2Fey-mainstreaming-esg-in-the-pre-ipo-process.pdf%3Fdownload/RK=2/RS=Z_ztHs_6jHZoJ0u9dzNOwEwT0ho-.

Fama, E. F. *et al.* (1992) *The Cross-Section of Expected Stock Returns*, *THE JOURNAL OF FINANCE **. doi: 10.1111/j.1540-6261.1992.tb04398.x.

Fama, E. F. and French, K. R. (1993) 'Common risk factors in the returns on stocks and bonds*', *Journal of Financial Economics*, 33, pp. 3–56.

Fama, E. and French, K. (1996) 'Multifactor Explanations of Asset Pricing Anomalies', *The Journal of finance (New York), 1996-03, Vol.51 (1), p.55-84*, 51(1), pp. 55–84.

Fama, F. and French, R. (1996) 'Multifactor Explanations of Asset Pricing Anomalies', LI(1).

Felix, N. (2020) 'A New Method for Regression Model Selection', *GSJ: Volume 8, Issue 12, December 2020, Online: ISSN 2320-9186*, 8(12), pp. 1858–1899. Available at: https://www.researchgate.net/publication/349915797_A_New_Method_for_Regression_Model_Selection/link/60473025a6fdcc9c7821a7be/download.

Feng, G. *et al.* (2020) 'Taming the Factor Zoo: A Test of New Factors', *THE JOURNAL OF FINANCE •*, LXXV(3). doi: 10.1111/jofi.12883.

Financial Times (2021) *'Green bubble' warnings grow as money pours into renewable stocks | Financial Times*. Available at: https://www.ft.com/content/0a3d0af8-7092-44c3-9c98-a513a22629be (Accessed: 25 April 2022).

Friedman, M. (1962) 'Capitalism and Freedom'. Available at: https://bibsys-almaprimo.hosted.exlibrisgroup.com/primo-explore/fulldisplay?docid=BIBSYS_ILS71472208830002201&context=L&vid=NTNU_UB&lang=no_NO&search_scope=default_scope&adaptor=Local Search Engine&isFrbr=true&tab=default_tab&query=any,contains,capitalism.

Ganatra, A., Panchal, G. and Kosta, Y. (2010) 'Searching Most Efficient Neural Network Architecture Using Akaike's Information Criterion (AIC)', *2010 International Journal of Computer Applications (0975 – 8887) Volume 1 – No. 5*. Available at: https://www.ijcaonline.org/journal/number5/pxc387242.pdf (Accessed: 18 April 2022).

Gradojevic, N. and Kukolj, D. (2022) 'Unlocking the black box: Non-parametric option pricing before and during COVID-19', *Annals of Operations Research*. doi: 10.1007/s10479-022-04578-7.

Gu, S., Kelly, B. and Xiu, D. (2018) 'NBER WORKING PAPER SERIES EMPIRICAL ASSET PRICING VIA MACHINE LEARNING Empirical Asset Pricing via Machine Learning'. Available at: http://www.nber.org/papers/w25398 (Accessed: 20 April 2022).

Gu, S., Kelly, B. and Xiu, D. (2020) 'Empirical Asset Pricing via Machine Learning *', *The Review of Financial Studies*, 33, pp. 2223–2273. doi: 10.1093/rfs/hhaa009.

Guo, H. (2004) 'Stock prices, firm size, and changes in the federal funds rate target', *Quarterly Review of Economics and Finance*, 44(4), pp. 487–507. doi: 10.1016/j.qref.2003.11.001.

Hlaváčková-Schindler, K. (2016) 'Prediction Consistency of Lasso Regression Does Not Need Normal Errors', *British Journal of Mathematics & Computer Science*, 19(4), pp. 1–7. doi: 10.9734/bjmcs/2016/29533.

Holthausen, R. W. and Zmijewski, M. E. (2012) 'Valuation with Market Multiples: How to Avoid Pitfalls When Identifying and Using Comparable Companies1', *Journal of Applied Corporate Finance*, 24(3), pp. 26–38. doi: 10.1111/J.1745-6622.2012.00387.X.

Hong, H. and Kacperczyk, M. (2009) 'The price of sin: The effects of social norms on markets', *Journal of Financial Economics*, 93(1), pp. 15–36. doi: 10.1016/J.JFINECO.2008.09.001.

Hou, K. *et al.* (2006) 'What Factors Drive Global Stock Returns?', *The Review of financial studies, 2011-08-01, Vol.24 (8), p.2527-2574*. Available at: https://www.jstor.org/stable/20869317?sid=primo&seq=1 (Accessed: 13 April 2022).

Hou, K. and Van Dijk, M. A. (2019) 'Resurrecting the size effect: Firm size, profitability shocks, and expected stock returns', *Review of Financial Studies*, 32(7), pp. 2850–2889. doi: 10.1093/rfs/hhy104.

Hu, S. (2012) 'Akaike Information Criterion Acoustic localisation of coronary artery stenosis View project Inverse Problems View project Akaike Information Criterion', *The American journal of medicine*. Available at: https://www.researchgate.net/publication/267201163 (Accessed: 18 April 2022).

Huang, Y. (2020) 'Lack of Robustness of Lasso and Group Lasso with Categorical Predictors: Impact of Coding Strategy on Variable Selection and Prediction', *Library Prize for Undergraduate Research, University of California Los Angeles*.

James, G., Hastie, T., *et al.* (2021) *An Introduction to Statistical Learning*. Second. Springer Texts In Statistics.

James, G., Witten, D., *et al.* (2021) 'Linear Model Selection and Regularization', pp. 225–288. doi: 10.1007/978-1-0716-1418-1_6.

Jegadeesh, N. and Titman, S. (1993) ‘Returns to Buying Winners and Selling Losers : Implications for Stock Market Efficiency’, *The Journal of Finance Vol. 48, No. 1 (Mar., 1993), pp. 65-91 (27 pages)*, 48(1), pp. 65–91. Available at: https://www.jstor.org/stable/2328882?seq=1.

Jensen, G. R. and Johnson, R. R. (1995) ‘Discount rate changes and security returns in the U.S., 1962-1991’, *Journal of Banking and Finance*, 19(1), pp. 79–95. doi: 10.1016/0378-4266(94)00048-8.

Kuha, J. (2004) ‘AIC and BIC: Comparisons of assumptions and performance’, *Sociological Methods and Research*, 33(2), pp. 188–229. doi: 10.1177/0049124103262065.

Kuhn, M. and Johnson, K. (2019) ‘Feature Engineering and Selection: A Practical Approach for Predictive Models: Chapter 5.7’, in. Chapman and Hall. Available at: https://bookdown.org/max/FES/.

Lewellen, J. (2014) ‘The Cross Section of Expected Stock Returns’, *SSRN Electronic Journal*, (August). doi: 10.2139/ssrn.2511246.

Liew, J. and Vassalou, M. (2000) ‘Can book-to-market, size and momentum be risk factors that predict economic growth?’, *Journal of Financial Economics*, 57(2), pp. 221–245. doi: 10.1016/S0304-405X(00)00056-8.

Lins, K. V *et al.* (2017) ‘Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis’, *The Journal of Finance*, 72(4), pp. 1785–1824. doi: 10.1111/JOFI.12505.

Lintner, J. (1965) ‘The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets’, *The Review of Economics and Statistics*, 47(1), pp. 13–37.

Long, J. T. *et al.* (2018) ‘Variation inflation factor-based regression modeling of anthropometric measures and temporal-spatial performance: Modeling approach and implications for clinical utility’, *Clinical Biomechanics*, 51(May 2017), pp. 51–57. doi: 10.1016/j.clinbiomech.2017.11.008.

Lundberg, S. M., Allen, P. G. and Lee, S.-I. (2017) ‘A Unified Approach to Interpreting Model Predictions’, *NIPS’17: Proceedings of the 31st International Conference on Neural Information Processing Systems*. Available at: https://github.com/slundberg/shap (Accessed: 20 April 2022).

Markowitz, H. (1952) ‘Portfolio Selection’, *The Journal of Finance*, 7(1), pp. 77–91. Available at: https://bibsys-almaprimo.hosted.exlibrisgroup.com/primo-explore/fulldisplay?docid=BIBSYS_ILS71524823940002201&context=L&vid=NTNU_UB&lang=no_NO&search_scope=default_scope&adaptor=Local Search Engine&isFrbr=true&tab=default_tab&query=any,contains,Portfolio Selection&offset=0 (Accessed: 29 April 2022).

Martani, D. and Khairurizka, R. (2009) ‘The effect of financial ratios, firm size, and cash flow from operating activities in the interim report to the stock return’, *Business Review*, 8(6), pp. 1537–1506. Available at: https://www.researchgate.net/publication/341671368_THE_EFFECT_OF_FINANCIAL_RATIOS_FIRM_SIZE_AND_OPERATING_CASH_FLOWS_ON_EARNING_PER_SHARE_EVIDENCE_FROM_LISTED_PLANTATION_COMPANIES_OF_COLOMBO_STOCK_EXCHANGE_CSE_IN_SRI_LANKA (Accessed: 12 April

2022).

McKinsey & Company (2018) *Why commercial use could be the future of carbon capture*. Available at: https://www.mckinsey.com/business-functions/sustainability/our-insights/why-commercial-use-could-be-the-future-of-carbon-capture (Accessed: 2 May 2022).

Merton, R. (1987) 'A Simple Model of Capital Market Equilibrium with Incomplete Information', *The Journal of Finance*, 42(3), pp. 483–510. doi: 10.1111/j.1540-6261.1987.tb04565.x.

Morningstar (2022) *Sustainable Funds Landscape - Highlights and Observations*. Available at: https://www.morningstar.com/articles/1080300/sustainable-funds-landscape-highlights-and-observations (Accessed: 11 June 2022).

Mossin, J. (1966) 'Equilibrium in a Capital Asset Market Author (s): Jan Mossin Reviewed work (s): Published by : The Econometric Society Stable URL : http://www.jstor.org/stable/1910098 .', *Econometrica*, 34(4), pp. 768–783.

Narinç, D., Öksüz Narinç, N. and Aygün, A. (2017) 'Growth curve analyses in poultry science', *World's Poultry Science Journal*, 73(2), pp. 395–408. doi: 10.1017/S0043933916001082.

NBIM (2021a) *Investeringsstrategien | Norges Bank Investment Management*. Available at: https://www.nbim.no/no/oljefondet/slik-er-fondet-investert/investeringsstrategien/ (Accessed: 12 May 2022).

NBIM (2021b) 'THE ASSET PRICING EFFECTS OF ESG INVESTING'. Available at: www.nbim.no (Accessed: 24 April 2022).

Nissim, D. (2019) 'EBITDA, EBITA, or EBIT?', *SSRN Electronic Journal*. doi: 10.2139/SSRN.2999675.

Pastor, L. *et al.* (2020) 'Sustainable Investing in Equilibrium Sustainable Investing in Equilibriumˇ Luboš', *Journal of Financial Economics Volume 142, Issue 2, November 2021, Pages 550-571*. doi: https://doi.org/10.1016/j.jfineco.2020.12.011.

Penman, S. (2013) 'Financial Statement Analysis and Security Evaluation'.

Penman, S. H., Richardson, S. A. and Tuna, I. (2007) 'The Book-to-Price Effect in Stock Returns: Accounting for Leverage', *Journal of Accounting Research*, 45(2), pp. 427–467. doi: 10.1111/J.1475-679X.2007.00240.X.

Petkova, R. and Zhang, L. (2005) 'Is value riskier than growth?', *Journal of Financial Economics*, 78(1), pp. 187–202. doi: 10.1016/j.jfineco.2004.12.001.

Rendleman, R. J., Jones, C. P. and Latané, H. A. (1982) 'Empirical anomalies based on unexpected earnings and the importance of risk adjustments', *Journal of Financial Economics*, 10(3), pp. 269–287. doi: 10.1016/0304-405X(82)90003-4.

Riis, C. and Moen, E. R. (2017) *Moderne Mikrøkonomi (Modern Microeconomics)*. 4th edn. Gyldendal.

Rosenberg, B. and Guy, J. (1995) 'Prediction of Beta from Investment Fundamentals', *Financial Analysts Journal*, 51(1), pp. 101–112. doi: 10.2469/faj.v51.n1.1866.

Ross, S. (1976) ‘Return, Risk and Arbitrage’, *Rodney L. White Center for Financial Research Working Papers 17-73, Wharton School Rodney L. White Center for Financial Research.* Available at: https://www.researchgate.net/publication/5071500_Return_Risk_and_Arbitrage.

Sanjar, K. *et al.* (2020) ‘Missing data imputation for geolocation-based price prediction using KNN-MCF method’, *ISPRS International Journal of Geo-Information*, 9(4). doi: 10.3390/ijgi9040227.

Sharpe, W. F. (1964) ‘CAPITAL ASSET PRICES: A THEORY OF MARKET EQUILIBRIUM UNDER CONDITIONS OF RISK*’, *The Journal of Finance*, 19(3), pp. 425–442. doi: 10.1111/J.1540-6261.1964.TB02865.X.

Sharpe, W. F. (1966) ‘Security Prices , Risk , and Maximal Gains from Diversification : Reply Author (s): W . F . Sharpe Source : The Journal of Finance , Dec ., 1966 , Vol . 21 , No . 4 (Dec ., 1966), pp . 743-744 Published by : Wiley for the American Finance Association’, 21(4), pp. 743–744.

Shefrin, H. M. and Statman, M. (1984) ‘Explaining investor preference for cash dividends’, *Journal of Financial Economics*, 13(2), pp. 253–282. doi: 10.1016/0304-405X(84)90025-4.

Shmueli, G. (2010) ‘To Explain or to Predict?’, *Statistical Science*, 25(3), pp. 289–310. doi: 10.1214/10-STS330.

Silgjerd, O. (2021) *ESG : All Bark and No Bite ?* Available at: https://openaccess.nhh.no/nhh-xmlui/handle/11250/2774583.

Sklearn (2022) *sklearn.model_selection.GridSearchCV — scikit-learn 1.1.1 documentation*. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html (Accessed: 15 June 2022).

Soliman, M. T. (2011) ‘The Use of Dupont Analysis by Market Participants’, *SSRN Electronic Journal*. doi: 10.2139/SSRN.1101981.

Statman, M., Fisher, K. L. and Anginer, D. (2008) ‘Affect in a behavioral asset-pricing model’, *Financial Analysts Journal*, 64(2), pp. 20–29. doi: 10.2469/faj.v64.n2.8.

Storebrand (2022) ‘Nøkkelinformasjon for investorer’, 1(2).

The Economist (2021) *The world’s biggest carbon-removal plant switches on*. Available at: https://www.economist.com/science-and-technology/2021/09/18/the-worlds-biggest-carbon-removal-plant-switches-on (Accessed: 2 May 2022).

Titman, S., Wei, K. C. J. and Xie, F. (2004) ‘Capital Investments and Stock Returns’, *Journal of financial and quantitative analysis, 2004-12, Vol.39 (4), p.677-700*, 39(4), pp. 677–700.

UNPRI (2020) *Annual Report 2020*. Available at: https://www.unpri.org/annual-report-2020/foreword (Accessed: 11 June 2022).

Uyar, A. (2016) ‘Evolution of Corporate Reporting and Emerging Trends’, *Journal of Corporate Accounting & Finance*, 27(4), pp. 27–30. doi: 10.1002/JCAF.22157.

Wang, H. and Sun, F. (2021) ‘Optimal Sensor Placement and Fault Diagnosis Model of PV Array of Photovoltaic Power Stations Based on XGBoost’, *IOP Conference Series: Earth and Environmental Science*, 661(1). doi: 10.1088/1755-1315/661/1/012025.

Weigand, A. (2019) ‘Machine learning in empirical asset pricing’, *Financial Markets and Portfolio Management*, 33(1), pp. 93–104. doi: 10.1007/s11408-019-00326-3.

Wooldridge, J. M. (2018a) ‘Introductory Econometrics: A modern approach, Chapter 2’, in. Cengage Learning, pp. 20–56.

Wooldridge, J. M. (2018b) ‘Introductory Econometrics’, in. Cengage Learning, pp. 40–42, 66–89.

Wooldridge, J. M. (2018c) ‘Introductory Econometrics Chapter 13’, in. Cengage Learning, pp. 226–251.

Zeidan, M. J. (2021) *Effects of Illegal Behavior on the Financial Performance of US Banking Institutions on JSTOR, Journal of Business Ethics*. Available at: https://www.jstor.org/stable/23327207 (Accessed: 25 April 2022).

Zuur, A. F., Ieno, E. N. and Elphick, C. S. (2010) ‘A protocol for data exploration to avoid common statistical problems’, *Methods in Ecology and Evolution*, 1(1), pp. 3–14. doi: 10.1111/j.2041-210x.2009.00001.x.

Appendices

A. Preprocessing

Year	Missing Observations
2012	27
2013	19
2014	18
2015	14
2016	11
2017	7
2018	4
2019	2
2020	1
2021	0

Table A-1. Companies in the S&P 500 index from 2012-2021 with unavailable data on returns as a consequence of de-listings, mergers and acquisitions. These observations was therefore omitted from the dataset.

Variable	2012-2016	2013-2017	2014-2018	2015-2019	2016-2020	2017-2021	Appearances	Average rank
Year	1	1	9	5	1	4	6	3,5
Relative Change Federal Funds Rate	2	2	6	9	2	1	6	3,7
Beta	5	5	7	7	3	3	6	5,0
Absolute Change Federal Funds Rate	3	6	20	1	4	2	6	6,0
Enterprise Value	8	4	2	6	9	13	6	7,0
Price-to-Free Cash Flow	0	8	1	2	5	7	5	4,6
Level Federal Funds Rate	4	9	11	0	6	11	5	8,2
Capital Expenditures	14	7	3	8	19	0	5	10,2
CO2 Equivalent Emissions Direct, Scope 1	19	11	8	15	20	0	5	14,6
Price-to-Cash Flow	13	0	0	4	7	9	4	8,3
Equity Risk Premium	6	14	5	18	0	0	4	10,8
Dividend	0	0	12	11	16	16	4	13,8
Price-to-Common Equity	13	0	0	17	12	15	4	14,3
Enterprise Value-to-EBITDA	10	0	0	14	17	18	4	14,8
Price-to-Book	0	0	16	16	18	12	4	15,5
Gross Profit Margin	17	17	0	0	15	14	4	15,8
Water Use-to-Revenues USD in Million	16	12	17	20	0	0	4	16,3
Price-to-Eearnings	0	0	0	12	11	8	3	10,3
Price-to-Operating Cash Flows	0	0	0	10	14	19	3	14,3
Price Momentum	11	0	0	0	0	6	2	8,5
Tax Rate - Actual	0	0	0	0	8	10	2	9,0
Full Time Employees	0	10	10	0	0	0	2	10,0
Income Available to Common Shareholders	7	15	0	0	0	0	2	11,0
EBIT Margin	12	13	0	0	0	0	2	12,5
EBITDA Margin	0	0	0	0	13	17	2	15,0
Energy Use Total	0	0	15	19	0	0	2	17,0
Dividend Yield to Common Stock Primary	18	18	0	0	0	0	2	18,0
Price-to-Free Operating Cashflows	0	19	19	0	0	0	2	19,0
Enterprise Value-to-EBIT	0	20	0	0	0	20	2	20,0
Return On Total Assets	0	0	14	0	0	0	1	14,0
Total Debt-to-EBITDA	15	0	0	0	0	0	1	15,0
Price-to-Dividends	0	16	0	0	0	0	1	16,0
CO2 Equivalent Emissions Total	0	0	18	0	0	0	1	18,0
Total CO2 Equivalent Emissions to Revenues USD in million	20	0	0	0	0	0	1	20,0

Table A-2.Variable ranking from 1-20 after importance determined by SHAP for the six time windows, and average ranking when included. Variables included less than two times marked by grey.

Variable	2017-2021		2016-2020		2015-2019		2014-2018		2013-2017		2012-2016		Appearances	
	UR	R	UR	R	UR	R	UR	R	UR	R	UR	R	UR	R
Equity Risk Premium	9,18	9,2	3,3	3,32	-2,04	-1,98	-1,71	1,67	-3,12	3,05	-0,73	-0,77	6	6
Enetrprise Value	0,25	0,24	1,08	1,1	1,77	1,6	0,87	1,12	0,37	0,57	0	0,37	6	5
Corporate Social Responsibility Strategy Score	0	0,005	0	0,014	0	0,003	0	0,01	0	0,016	0	-0,03	6	0
Dividend	-2,57	-2,37	-2,8	2,79	-2,3	-2,25	-1,58	-1,45	-1,55	-1,41	0,13	0	5	6
Relative Change Federal Funds Rate	-17,45	-17,5	0	0	3,33	3,45	-0,16	-0,1	-5,83	-5,87	-3,97	-3,9	5	5
Tax Rate - Actual	1,08	0,99	0,36	0,37	-0,08	-0,07	0	0	0,08	-0,05	-0,07	0,04	5	5
Beta	4,63	5,22	5,3	5,17	-1,52	-1,66	1,1	-1,2	0	0	0	-0,15	5	4
Year	-22	-22,1	14,7	14,73	0	0	0,1	0,067	0	0	-4,98	-4,95	4	4
Community Score	0	-0,02	0	0	0	0	0	0,026	0	0,05	0	0,02	4	0
Total Debt-to-EBITDA	0,52	0,56	-0,09	-0,07	0	0	-0,12	-0,16	0	0	0,1	0	3	4
Enterprise Value-to-EBIT	0	0	0	0	0,17	0,177	0	0	0,19	0,2	0,43	0,41	3	3
Level Federal Funds Rate	-23,66	23,6	-29,6	-29,6	0	0	0	0	0	0	0	5,28	3	2
Price-to-Earnings	-0,04	0	0	0	0	0	0,29	0,27	0,07	0,04	0	0	2	3
Gross Profit Margin	0	0	0	0	0,79	0	0	0	-0,05	-0,1	-0,31	-0,3	2	3
Absolute Change Federal Funds Rate	0	0	21,8	21,79	3,7	3,6	0	0	0	0	0	0	2	2
Long Term Debt-to-Total Capital	0	0	0	0	0,38	0,36	0,44	0,41	0	0	0	0	2	2
Price-to-Free-Cash-Flow	2,28	2,01	2,04	2,1	0	0	0	0	0	0	0	0	2	2
Return on Total Assets	0	0	0	0	0,51	0,53	0,004	0	0	0	0	-0,2	2	2
Product Responsibility Score	0	0,25	0	0,004	0	0	0	0	0	0	0	0	2	0
Total CO2 Equivalent Emissions to Revenues USD	0	0,45	0	0	0	0	0	0	0	0	0	-0,09	2	0
Price-to-Book	-0,008	0	-0,28	-0,3	0	0	0	0	0	0	0	0	1	2
Net Debt-to-Enterprise Value	0	0	0	0	0,8	-0,88	-0,04	0	0	0	0	0	1	2
Return On Net Operating Assets	0,15	0,056	0	0	0	0	0	0	0	0	0	0	1	1
Price-to-Cash Flow	-0,62	-0,58	0	0	0	0	0	0	0	0	0	0	1	1
EBITDA Margin	0,56	0,62	0	0	0	0	0	0	0	0	0	0	1	1
Dividend Yield to Common Stock Primary	0	0,041	0	0	0	0	0	0	0	0	0	0	1	0
Effective Tax Rate	0	0	0	0,53	0	0	0	0	0	0	0	0	1	0
Price-to-Dividend	0	0	0	0	0	0	0	0,09	0	0	0	0	1	0
ESG Score	0	0,065	0	0	0	0	0	0	0	0	0	0	1	0
Enterprise Value-to-EBITDA	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Table A-3. Coefficients from Lasso Regression. Variables included in less than two windows marked by grey.

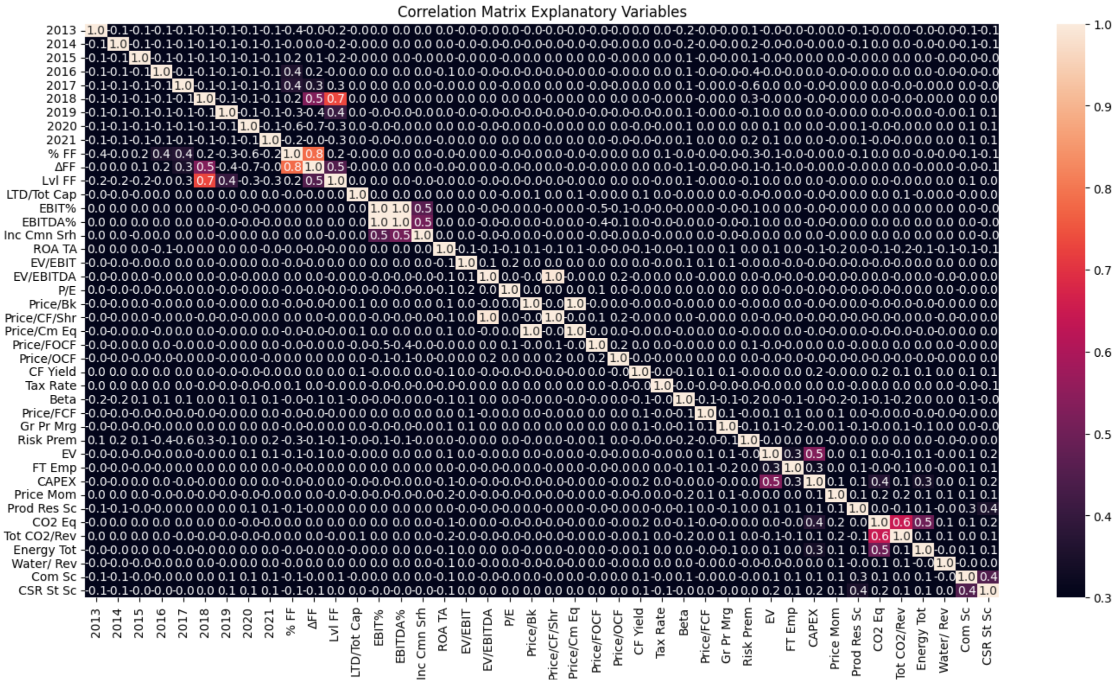

Figure A-1.Correlation between explanatory variables selected from XGBoost and Lasso regression. All bellow 0,3 set to zero.

Correlation Pair	Variables		Correlation Coefficient
1	Relative Change Federal Funds Rate	Absolute Change Federal Funds Rate	0,8
2	EBIT Margin	EBITDA Margin	1
3	Enterprise Value-to-EBITDA	Price-to-Cashflow	1
4	Price-to-Book	Price-to-Common Equity	1

Table A-4. Pairwise correlation between explanatory variables with VIF scores over 3,5.

Variable	VIF Factor
Absolute Change Federal Funds Rate	164 343 957 062,80
Level Federal Funds Rate	10 370 076 141,57
Relative Change Federal Funds Rate	1 795 579 796,72
EBITDA Margin	75 666,80
EBIT Margin	63 636,52
Price-to-Book	16 957,92
Price-to-Common Equity	16 957,29
Income Available to Common Shareholders	568,32
Price-to-Cash Flow	452,77
Enterprise Value-to-EBITDA	452,56
CO2 Equivalent Emissions Direct, Scope 1	2,92
Equity Risk Premium	2,79
Total CO2 Equivalent Emissions To Revenues USD in Million	2,19
Capital Expenditures	2,12
Price-to-Operating Cash Flow	2,09
Energy Use Total	2,00
Enterprise Value	1,60
Price-to-Free Operating Cash Flow	1,54
Beta	1,49
Corporate Social Responsibility Strategy Score	1,45
Gross Profit Margin	1,43
Full-Time Employees	1,35
Community Score	1,34
Product Responsibility Score	1,27
Return On Total Assets	1,19
Price Momentum	1,19
Water Use To Revenues USD in Million	1,13
Dividend	1,13
Long Term Debt to Total Capital	1,08
Enterprise Value-to-EBIT	1,07
Historic Price-to-Earnings	1,06
Price-to-Free Cash Flow	1,05
Tax Rate - Actual	1,01

Table A-5. VIF factors of explanatory variables in decreasing order. Variables with VIF higher than cutoff of 3,5 is marked by grey. Dark grey is variables that was removed due to multicollinearity.

First Generation Varaible Set (G1)	
Unrestrcited varialbe selection (G1UR)	**Restricted Variable Selection (G1R)**
Carbon Intensity per Energy Produced	Absolute change Federal Funds Rate
CO2 Equivalent Emissions Direct, Scope 1	Beta
CO2 Equivalent Emissions Indirect, Scope 2	Capital Expenditures
CO2 Equivalent Emissions Indirect, Scope 3	Dividend
CO2 Equivalent Emissions Indirect, Scope 3 To Revenues USD in Million	Dividend Yield Common Stock Primary
CO2 Equivalent Emissions Total	EBIT Margin
Community Score	EBITDA Margin
Corporate Social Responsibility Strategy Score	Effective Tax Rate
Downstream scope 3 emissions End-of-life Treatment of Sold Products (M/NM)	Enterprise Value
Downstream scope 3 emissions Investments (M/NM)	Enterprise Value/EBIT
Downstream scope 3 emissions Processing of Sold Products (M/NM)	Equity Risk Premium
Downstream scope 3 emissions Transportation and Distribution (M/NM)	Full-Time Employees
Downstream scope 3 emissions Use of Sold Products (M/NM)	Gross Profit Margin
Energy Use Total	Historic Enterprise Value/EBITDA
Environment Management Team (T/F)	Income Available to Common Shareholders
Environment Management Training (T/F)	Level Federal Funds Rate
Environmental Innovation Score	Long Term Debt to Total Capital
Environmental Materials Sourcing (T/F)	Net Debt-to-Enterprise Value
Environmental Pillar Score	Price Momentum
ESG Score	Price-to-Book
GHG Emissions Method (Binary - Missing/Not Missing)	Price-to-Cash Flow
NOx Emissions (M/NM)	Price-to-Common Equity
Policy Emissions Score	Price-to-Dividends
Policy Energy Efficiency (T/F/M)	Price-to-Earnings
Policy Environmental Supply Chain (T/F/M)	Price-to-Free Cash Flow
Policy Sustainable Packaging (T/F/M)	Price-to-Free Operating Cash Flow
Policy Water Efficiency (T/F/M)	Price-to-Operating Cash Flow
Product Responsibility Score	Relative Change Federal Funds Rate
Renewable Energy Use Ratio	Return On Total Assets
Renewable Energy Use Ratio Score	Revenue
Resource Reduction Policy (T/F/M)	Tax Rate - Actual
Resource Reduction Targets (T/F/M)	Tot Debt Cap-to-EBITDA
Resource Use Score	Total Debt to Total Equity
Return on Net Operating Assets - Mean	Year
SDG 7 Affordable and Clean Energy (T/F/M)	
SOx Emissions (M/NM)	
Targets Emissions Score	
Targets Energy Efficiency (T/F/M)	
Targets Water Efficiency (T/F/M)	
Total CO2 Equivalent Emissions-to-Revenues USD in Million	
Total Energy Use-to-Revenues USD In Million	
Total Renewable Energy	
Toxic Chemicals Reduction (T/F/M)	
Upstream scope 3 emissions Business Travel (M/NM)	
Upstream scope 3 emissions Capital goods (M/NM)	
Upstream scope 3 emissions Fuel- and Energy-related Activities (M/NM)	
Upstream scope 3 emissions Purchased goods and services (M/NM)	
Upstream scope 3 emissions Transportation and Distribution (M/NM)	
Upstream scope 3 emissions Waste Generated in Operations (M/NM)	
VOC or Particulate Matter Emissions Reduction Score	
Waste Recycled To Total Waste (M/NM)	
Water Use-to-Revenues USD in Million	
Absolute change Federal Funds Rate	
Beta	
Capital Expenditures	
Dividend	
Dividend Yield Common Stock Primary	
EBIT Margin	
EBITDA Margin	
Effective Tax Rate	
Enterprise Value	
Enterprise Value-to-EBIT	
Equity Risk Premium	
Full-Time Employees	
Gross Profit Margin	
Enterprise Value-to-EBITDA	
Income Available to Common Shareholders	
Level Federal Funds Rate	
Long Term Debt to Total Capital	
Net Debt-to-Enterprise Value	
Price Momentum	
Price-to-Book	
Price-to-Cash Flow	
Price-to-Common Equity	
Price-to-Dividends	
Price-to-Earnings	
Price-to-Free Cash Flow	
Price-to-Free Operating Cash Flow	
Price-to-Operating Cash Flow	
Relative Change Federal Funds Rate	
Return On Total Assets	
Revenue	
Tax Rate - Actual	
Tot Debt Cap-to-EBITDA	
Total Debt to Total Equity	
Year	

Table A-6. Variables included in first-generation variable selection (G1) with unrestricted (UR) and restricted (R) sub-selection. Binary variables representing missing and non-missing observation indicated by (M/NM) and True/False statements indicated by (T/F)

Second Generation Variable Set (G2)	
Unrestrcited varialbe selection (G2UR)	**Restricted Variable Selection (G2R)**
CO2 Equivalent Emissions Direct, Scope 1	Beta
Community Score	Capital Expenditures
Corporate Social Responsibility Strategy Score	Dividend %
Energy Use Total	EBIT Margin
Product Responsibility Score	Enterprise Value
Total CO2 Equivalent Emissions To Revenues USD in Million	Enterprise Value-to-EBIT
Water Use To Revenues USD in Million	Full-Time Employees
Beta	Gross Profit Margin
Capital Expenditures	Price-to-Book
Dividend %	Price-to-Free Operating Cash Flow
EBIT Margin	Income Available to Common Shareholders
Enterprise Value	Level Federal Funds Rate
Enterprise Value-to-EBIT	Long Term Debt to Total Capital
Full-Time Employees	Price Momentum
Gross Profit Margin	Price-to-Free Cash Flow
Price-to-Book	Price-to-Free Operating Cash Flow
Price-to-Free Operating Cash Flow	Price-to-Cash Flow
Income Available to Common Shareholders	Relative Change Federal Funds Rate
Level Federal Funds Rate	Return On Total Assets
Long Term Debt to Total Capital	Tax Rate - Actual
Price Momentum	Equity Risk Premium
Price-to-Free Cash Flow	Year
Price-to-Operating Cash Flow	
Price-to-Cash Flow	
Relative Change Federal Funds Rate	
Return On Total Assets	
Tax Rate - Actual	
Equity Risk Premium	
Year	

Table A-7. Variables included in second-generation variable selection (G2) with unrestricted (UR) and restricted (R) sub-selection.

B. Goodness-of-fit

Time Window	SST
2012-2016	1527090
2013-2017	1504912
2014-2018	1364614
2015-2019	1605992
2016-2020	3972536
2017-2021	4141055

Table B-1. Total sum of squares (SST) for each time window

Out-of-sample R^2: G2	XGBoost		OLS	
	UR	R	UR	R
W1	-0,625	-0,760	-0,020	-0,020
W2	-0,009	-0,010	-0,054	-0,056
W3	-0,004	-0,038	-2,870	-0,076
W4	-0,340	-0,357	-0,447	-0,470
W5	-2,210	-2,230	-0,105	-0,100
W6	-7,400	-7,240	-0,890	-0,091

Table B-2. Showing Out-of-sample R^2 for XGBoost and OLS models using for UR and R using G2.

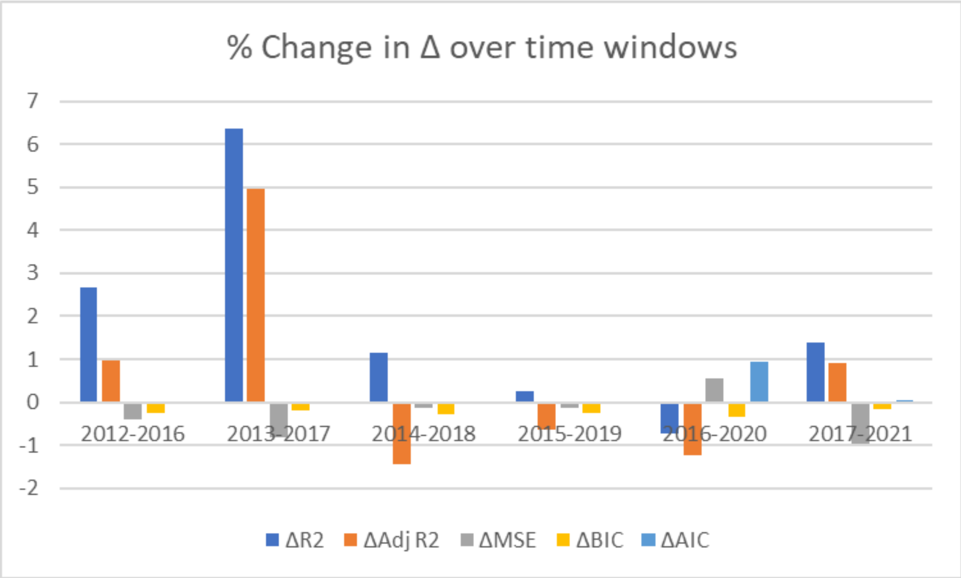

Figure B-1. The % change in delta over the six time periods based on data for the XGB regression. Showing much variation in all criteria over time, but not following any immediately recognizable pattern.

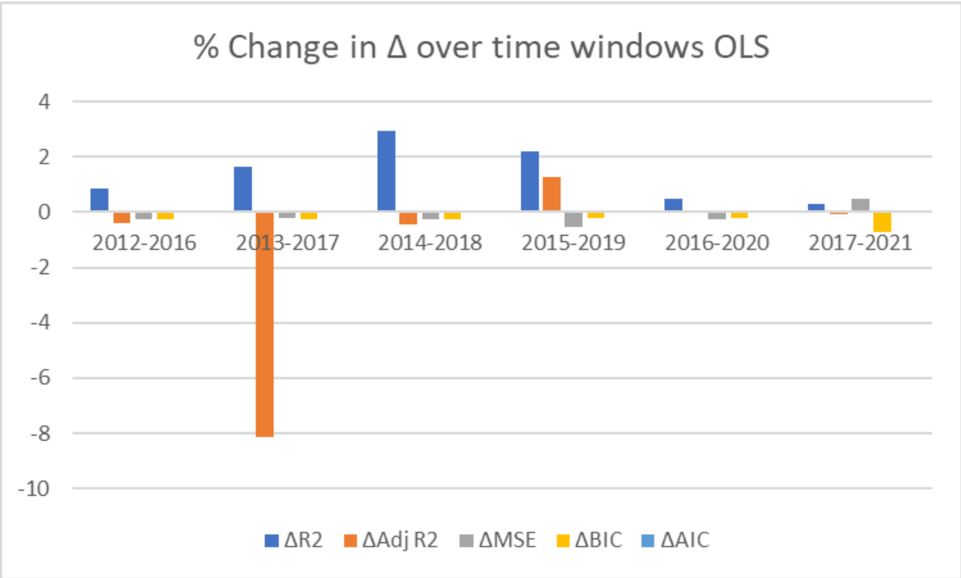

Figure B-2. The % change in delta over the six time periods based on data for the OLS regression. Showing much variation in all criteria over time, but not following any immediately recognizable pattern.

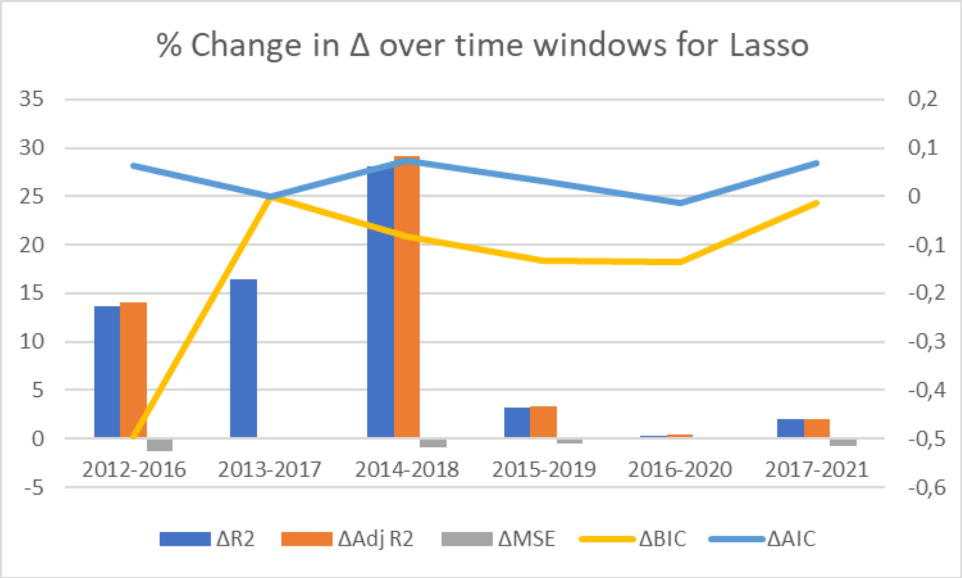

Figure B-3. The % change in delta over the six time periods based on data from the Lasso regression. Showing much variation in all criteria over time, but not following any immediately recognizable pattern.

C. Variable Importance

OLS	2012-2016		2013-2017		2014-2018		2015-2019		2016-2020		2017-2021		
Variable Set: G2	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Significant windows
Relative Change Federal Funds Rate	0,08000000000	0,01300000000	-0,04800000000	0,03200000000	0,16000000000	0,02000000000	0,20000000000	0,00000000000	-0,08000000000	0,00000000000	-0,10000000000	0,00000000000	*6*
Dividend	-0,94000000000	0,02900000000	-2,04000000000	0,00000000000	-2,11000000000	0,00000000000	-2,95000000000	0,00000000000	-3,08000000000	0,00000000000	-2,71000000000	0,00000000000	*6*
Enterprise Value	0,00000000003	0,00000000000	0,00000000003	0,00000000000	0,00000000003	0,00000000000	0,00000000004	0,00000000000	0,00000000002	0,00000000000	0,00000000001	0,01000000000	*6*
Beta	-0,47000000000	0,72000000000	1,32000000000	0,28000000000	-2,78000000000	0,01400000000	-4,90000000000	0,00000000000	12,20000000000	0,00000000000	11,90000000000	0,00000000000	*4*
Capital Expenditures	-0,00000000080	0,00100000000	-0,00000000070	0,00100000000	-0,00000000070	0,00100000000	-0,00000000059	0,00600000000	-0,00000000005	0,08000000000	-0,00000000050	0,10000000000	*4*
Constant	21,50000000000	0,02400000000	17,75000000000	0,06700000000	12,50000000000	0,03000000000	9,10000000000	0,19000000000	24,20000000000	0,02800000000	3,57000000000	0,70000000000	*3*
2015	-17,94000000000	0,00000000000	-6,40000000000	0,01400000000	-6,64000000000	0,00100000000							*3*
2016	-3,50000000000	0,07500000000	11,50000000000	0,00000000000	0,80000000000	0,75000000000	5,80000000000	0,13000000000					*3*
2017			0,76000000000	0,32000000000	-4,40000000000	0,10000000000	-7,60000000000	0,17000000000	1,40000000000	0,60000000000			*3*
Level Federal Funds Rate	-0,66000000000	0,43000000000	4,60000000000	0,00000000000	2,10000000000	0,02600000000	7,35000000000	0,03700000000	-7,15000000000	0,06200000000	-5,15000000000	0,10000000000	*3*
Return on Total Assets	-0,18000000000	0,01800000000	0,02300000000	0,75000000000	0,14000000000	0,04400000000	0,25000000000	0,00100000000	0,03800000000	0,68000000000	-0,02000000000	0,83000000000	*3*
Price Momentum	-0,04700000000	0,01100000000	-0,02000000000	0,24000000000	-0,00200000000	0,91000000000	0,00030000000	0,98000000000	-0,05000000000	0,02700000000	0,10000000000	0,00000000000	*3*
2013	18,14000000000	0,00100000000											*2*
2018					0,22000000000	0,83000000000	-0,60000000000	0,60000000000	7,80000000000	0,00000000000	3,50000000000	0,01500000000	*2*
2019							9,30000000000	0,11000000000	-23,90000000000	0,00000000000	-23,50000000000	0,00000000000	*2*
2020									31,00000000000	0,00000000000	35,30000000000	0,00000000000	*2*
Long Term Debt-to-Total Capital	0,02400000000	0,10800000000	0,02300000000	0,12000000000	0,04000000000	0,00600000000	0,03000000000	0,04000000000	0,02000000000	0,29000000000	0,02000000000	0,28000000000	*2*
Tax Rate - Actual	-0,01100000000	0,12000000000	-0,01000000000	0,13000000000	-0,01000000000	0,16000000000	-0,01000000000	0,10000000000	0,07000000000	0,04800000000	0,07000000000	0,03000000000	*2*
Price-to-Free Cash Flows	0,09000000000	0,05200000000	0,00530000000	0,23000000000	0,00730000000	0,08900000000	0,00700000000	0,14000000000	0,02700000000	0,00000000000	0,02000000000	0,00000000000	*2*
Gross Profit Margin	-0,05600000000	0,03500000000	-0,04000000000	0,08000000000	-14,10000000000	0,00000000000	0,06000000000	0,02400000000	-0,09000000000	0,79000000000	-0,00700000000	0,84000000000	*2*
2014	1,78000000000	0,57000000000	0,32000000000	0,86000000000									*1*
2021											-25,00000000000	0,00000000000	*1*
EBIT Margin	-0,00020000000	0,74000000000	-0,01500000000	0,02400000000	0,00400000000	0,38000000000	0,00300000000	0,50000000000	-0,00004000000	0,49000000000	-0,00004000000	0,33000000000	*1*
Enterprise Value-to-EBIT	0,00660000000	0,03200000000	0,00600000000	0,07100000000	0,00320000000	0,30000000000	0,00600000000	0,05600000000	0,00600000000	0,14000000000	0,00500000000	0,54000000000	*1*
Price-to-Earnings	0,00260000000	0,33800000000	0,00400000000	0,14000000000	0,01400000000	0,06400000000	0,00200000000	0,76000000000	-0,00900000000	0,34000000000	-0,01000000000	0,05000000000	*1*
Price-to-Book	0,00360000000	0,79000000000	0,00520000000	0,70000000000	-0,01000000000	0,50000000000	-0,02300000000	0,12000000000	-0,05000000000	0,01700000000	-0,04000000000	0,05200000000	*1*
CO2 Equivalent Emissions Direct, Scope 1	0,00000000200	0,97000000000	-0,00000000180	0,97000000000	0,00000006000	0,24000000000	0,00000080000	0,12000000000	0,00000014000	0,04500000000	0,00000011000	0,07000000000	*1*
Energy Use Total	0,00000000150	0,82000000000	-0,00000000040	0,95000000000	-0,00000000900	0,10000000000	-0,00000012000	0,01300000000	-0,00000000700	0,19000000000	-0,00000000200	0,67000000000	*1*
Water Use-to-Revenues USD in Million	-0,00000400000	0,03300000000	-0,00000140000	0,25000000000	-0,00000090000	0,37000000000	-0,00000140000	0,12000000000	-0,00000020000	0,80000000000	0,00000020000	0,77000000000	*1*
Income Available to Common Shareholders	0,00004000000	0,53000000000	0,00010000000	0,89000000000	-0,00030000000	0,40000000000	-0,00020000000	0,50000000000	0,00000600000	0,37000000000	0,00000600000	0,16000000000	*0*
Price-to-Free Cash Flows	-0,01100000000	0,40000000000	-0,00080000000	0,46000000000	-0,00040000000	0,73000000000	-0,00080000000	0,47000000000	-0,00100000000	0,56000000000	-0,00300000000	0,11000000000	*0*
Price-to-Free Operating Cash Flow	0,00540000000	0,40000000000	0,00900000000	0,17000000000	0,00080000000	0,86000000000	-0,00300000000	0,49000000000	-0,00260000000	0,67000000000	-0,00500000000	0,35000000000	*0*
Price-to-Operating Cash Flow	0,01000000000	0,47700000000	-0,00520000000	0,64000000000	0,00080000000	0,34000000000	0,00100000000	0,23000000000	0,00060000000	0,62000000000	0,00090000000	0,40000000000	*0*
Equity Risk Premium	-1,05000000000	0,62000000000	-0,92000000000	0,67000000000	-3,66000000000	0,11000000000	-5,10000000000	0,05200000000	-0,85000000000	0,77000000000	0,74000000000	0,74000000000	*0*
Full-Time Employees	-0,00000500000	0,28000000000	-0,00000500000	0,24000000000	-0,00000350000	0,41000000000	-0,00000350000	0,42000000000	-0,00000400000	0,50000000000	-0,00000100000	0,83000000000	*0*
Product Responsibility Score	0,00400000000	0,85000000000	0,00040000000	0,99000000000	0,02000000000	0,30000000000	0,02400000000	0,23000000000	0,04000000000	0,13000000000	0,04000000000	0,13000000000	*0*
Community Score	0,00230000000	0,41700000000	0,05000000000	0,07000000000	0,02800000000	0,33000000000	0,04000000000	0,19000000000	-0,01400000000	0,73000000000	-0,02800000000	0,50000000000	*0*
Corporate Social Responsibility Strategy Score	-0,00660000000	0,76000000000	-0,01600000000	0,44000000000	-0,01000000000	0,54000000000	0,00100000000	0,96000000000	-0,01600000000	0,58000000000	-0,01000000000	0,69000000000	*0*

Table C-1. Estimated regression coefficients and associated p-values for selected variables. Darkest grey indicating the time windows the various dummy variables was not included. Light grey marks p-values at minimum 5% significance level. Variables significant in less than two time windows (Except *"2014"* and *"2021")* in medium grey.

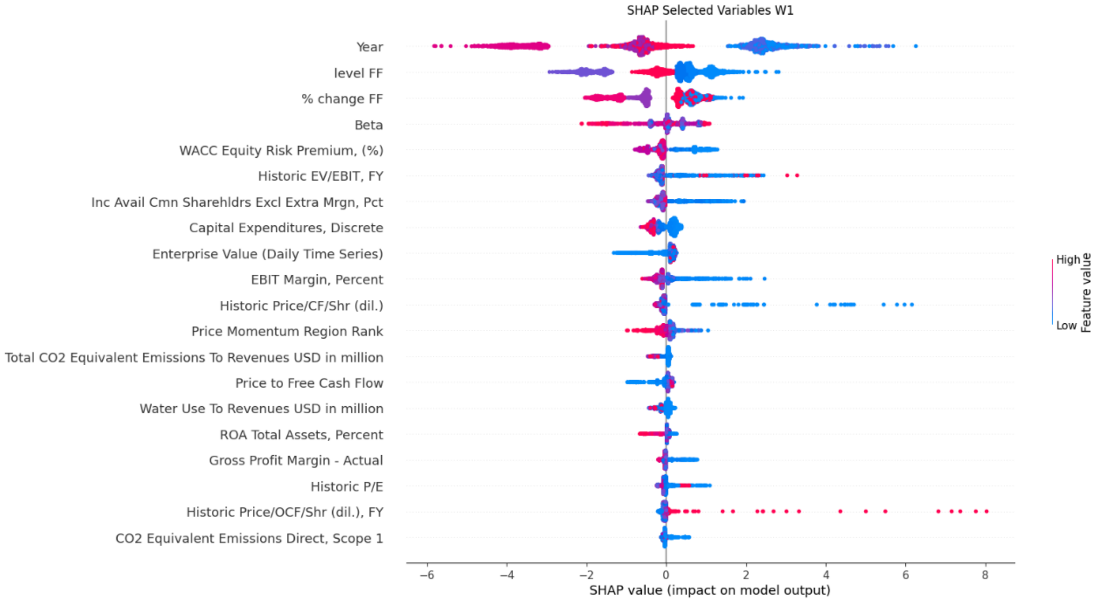

Figure C-1.SHAP for G2 XGBoost regression W1. Each dot indicates an individual observation. Red indicates high observed values for the respective variables, and blue low ones. Strong color indicates more extreme high or low values. The horizontal axis has the same unit as the response variable, namely percentage increase in total returns during the past 52 weeks. A red dot in the left-hand side, as seen for *"Year"*. Indicates that observations for high value, meaning more recent years, have negative impact on XGBoost's predicted expected returns. While the blue dots on the right hand side indicate that years further back in time indicate the opposite.

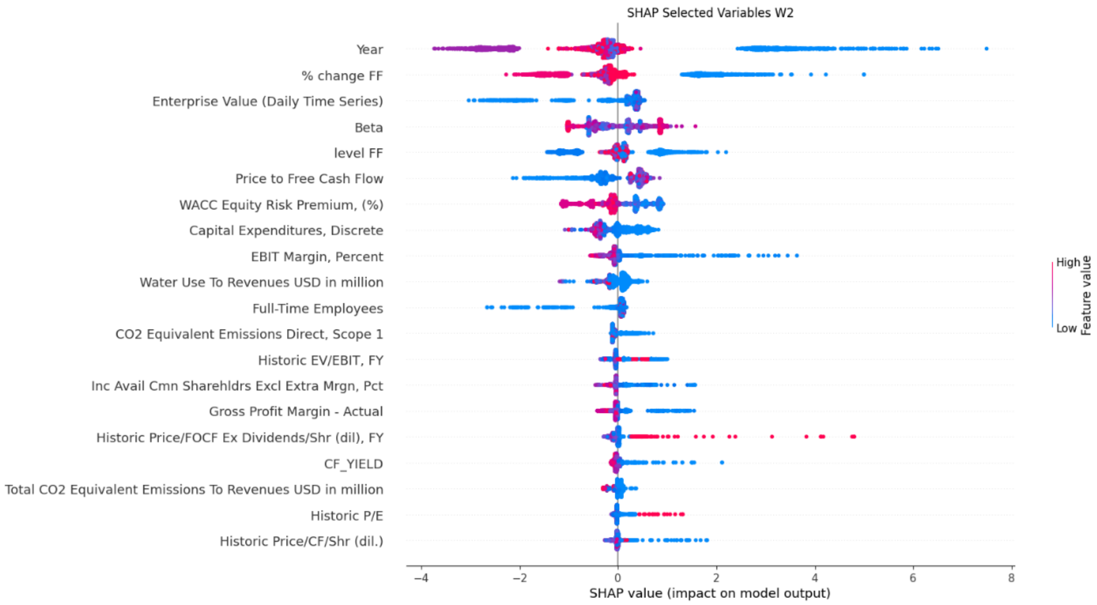

Figure C-2. SHAP for 2.gen XGBoost regression W2. Each dot indicates an individual observation. Red indicates high observed values for the respective variables, and blue low ones. Strong color indicates more extreme high or low

values. The horizontal axis has the same unit as the response variable, namely percentage increase in total returns during the past 52 weeks

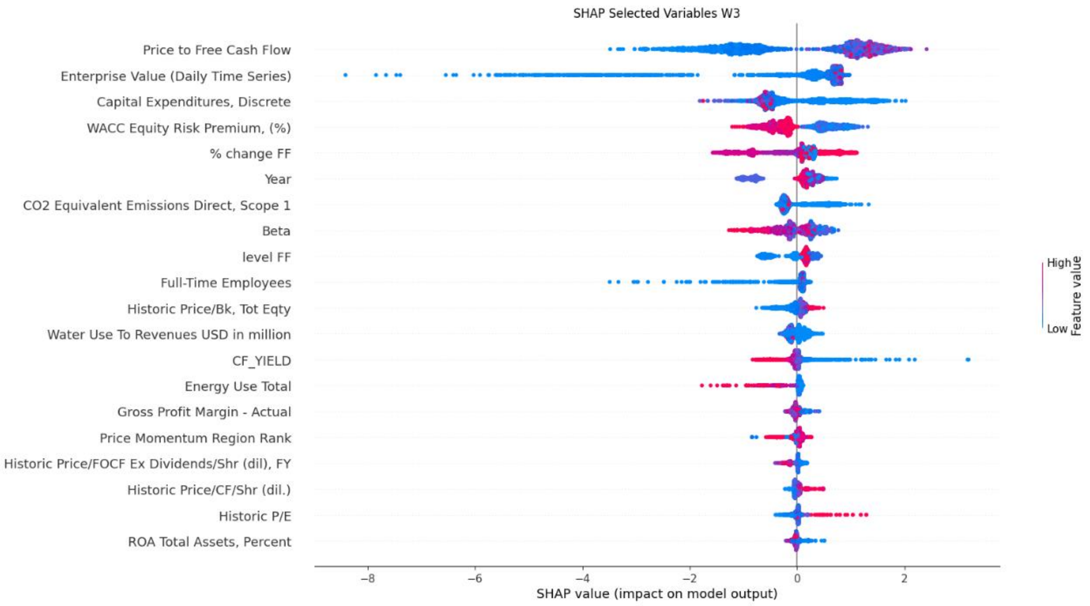

Figure C-3. SHAP for 2.gen XGBoost regression W3. Each dot indicates an individual observation. Red indicates high observed values for the respective variables, and blue low ones. Strong color indicates more extreme high or low values. The horizontal axis has the same unit as the response variable, namely percentage increase in total returns during the past 52 weeks

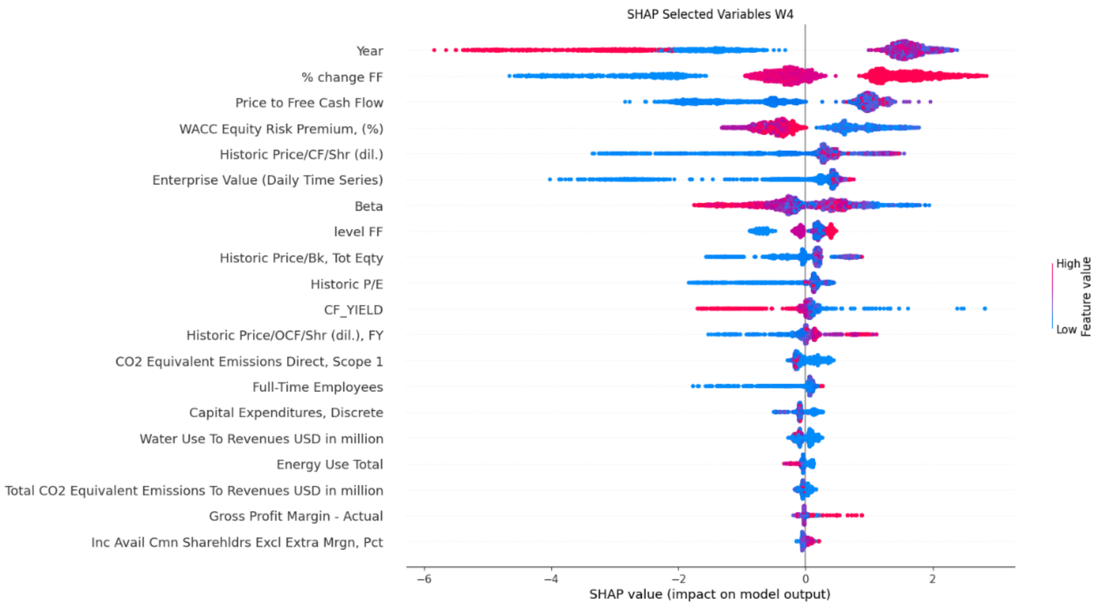

Figure C-4. SHAP for 2.gen XGBoost regression W4. Each dot indicates an individual observation. Red indicates high observed values for the respective variables, and blue low ones. Strong color indicates more extreme high or low values. The horizontal axis has the same unit as the response variable, namely percentage increase in total returns during the past 52 weeks

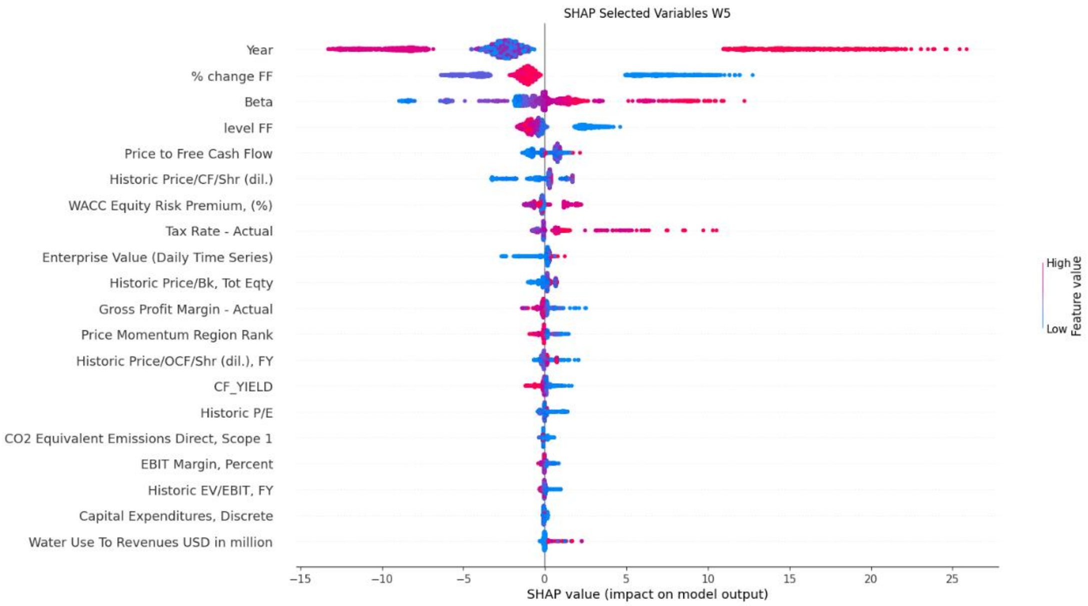

Figure C-5. SHAP for 2.gen XGBoost regression W5. Each dot indicates an individual observation. Red indicates high observed values for the respective variables, and blue low ones. Strong color indicates more extreme high or low values. The horizontal axis has the same unit as the response variable, namely percentage increase in total returns during the past 52 weeks

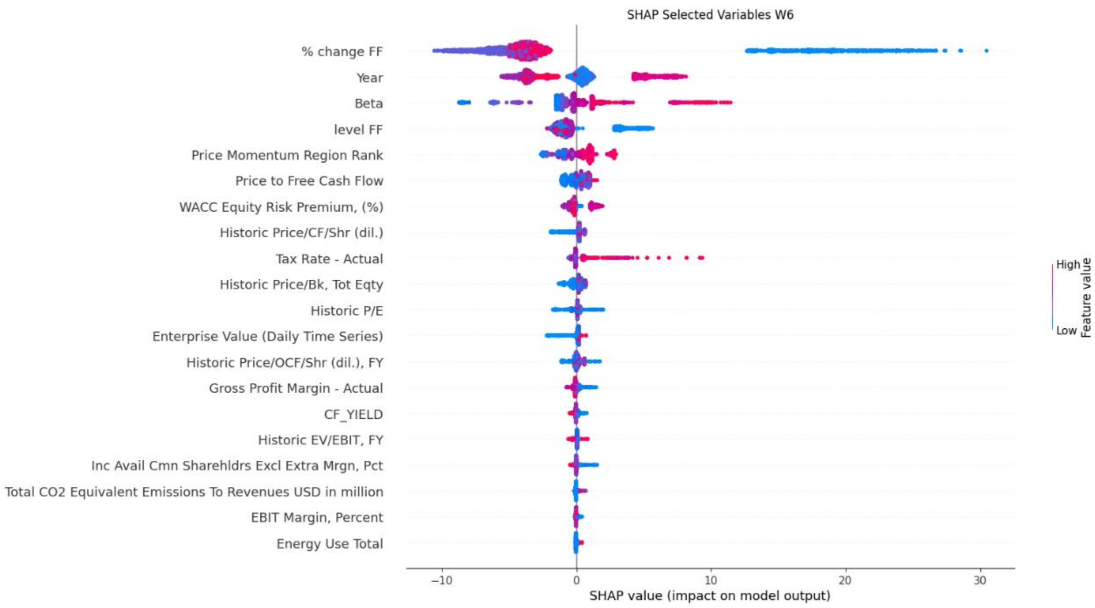

Figure C-6. SHAP for 2.gen XGBoost regression W6. Each dot indicates an individual observation. Red indicates high observed values for the respective variables, and blue low ones. Strong color indicates more extreme high or low values. The horizontal axis has the same unit as the response variable, namely percentage increase in total returns during the past 52 weeks

NTNU

Kunnskap for en bedre verden