

Tail-Risk and Sustainability: Can ESG scores accurately predict Value at Risk? – A machine learning based approach

Ulrich Mohme

Dissertation written under the supervision of Paul Karehnke

Dissertation submitted in partial fulfilment of requirements for the International MSc in Finance, at Universidade Católica Portuguesa and for the Master in Management at ESCP Business School, 20.05.2024.

Table of Contents

1. Introduction	4
2. Literature Review	9
2.1. <i>Importance of ESG Factors</i>	9
2.2. <i>ESG Factors, Firm Valuation and Stock Returns</i>	9
2.3. <i>Executive Behaviour towards ESG Risk</i>	11
2.4. <i>Link: Carbon, Climate and Financial Risk</i>	12
2.5. <i>ESG, Volatility and Tail Risk</i>	13
3. Data	15
4. Methodology	18
5. Results	22
5.1. <i>Descriptive Statistics</i>	22
5.2. <i>Descriptive Chart Analysis</i>	26
5.3. <i>Correlation Analysis</i>	30
5.4. <i>Findings from machine learning based predictions</i>	32
6. Conclusion	34
Publication bibliography	36
Appendix	39

List of Tables

Table 1: Descriptive statistics of all variables, cleaned from null variables but not from outliers.....	23
Table 2: Results from machine learning based predictions on VaR at the 1% and 5% confidence level	32

List of Figures

Figure 1: Histograms of the distributions with a bin size of 50.....	26
Figure 2: Density Plot of VaR at 1% and VaR at 5% stacked.....	27
Figure 3: Density plots of the distributions of the variables	28
Figure 4: Box plot of the distributions of all observe variables.....	29
Figure 5: Correlation Matrix of the analysed variables in style of a heat map.....	30
Figure 6: Regression plot of the ESG related scores and VaR at 1% confidence level.....	31
Figure 7: Regression plot of the ESG related scores and VaR at 5% confidence level.....	32

List of Equations

Equation 1: Return Index excluding dividends.....	16
Equation 2: Return Index including dividends	16

Title: Tail-Risk and Sustainability: Can ESG scores accurately predict Value at Risk? - A machine learning based approach

Author: Ulrich Mohme

Key Words: Tail Risk, Value at Risk, Sustainability, ESG, Machine Learning

Abstract English

I, Ulrich Mohme, find in the scope of the master thesis: “Tail-Risk and Sustainability: Can ESG scores accurately predict Value at Risk? – A machine learning based approach“ that predicting Value at Risk at the 1% and 5% confidence level by applying various machine learning algorithms onto ESG scores show low degrees of accuracy. Random Forest Regressors show the highest degree of accuracy from the algorithms used and the ESG as well as Environmental Score correlate most strongly with Value at Risk indicating the most significant predictive power. Data from companies listed in the S&P500 are used from the year 2000 to 2024. The findings imply ESG scores alone not to be a reliable predictor of Value at Risk at various significance levels. Yet a slightly linear correlation is detected and machine learning algorithms outperform benchmark linear regression models.

Abstract Portuguese

Eu, Ulrich Mohme, no âmbito da tese de mestrado: “Tail-Risk and Sustainability: Can ESG scores accurately predict Value at Risk? – A machine learning based approach“ encontro que prever o Valor em Risco (Value at Risk) nos níveis de confiança de 1% e 5%, aplicando vários algoritmos de aprendizado de máquina nas pontuações ESG (ambientais, sociais e de governança), mostra baixos graus de precisão. Os modelos de Regressão de Floresta Aleatória apresentam o maior grau de precisão entre os algoritmos utilizados, e as pontuações ESG, bem como a Pontuação Ambiental, correlacionam-se mais fortemente com o Valor em Risco, indicando o poder preditivo mais significativo. Foram utilizados dados de empresas listadas no S&P500 do ano 2000 a 2024. Os resultados implicam que as pontuações ESG sozinhas não são um indicador confiável do Valor em Risco em vários níveis de significância. No entanto, foi detectada uma ligeira correlação linear e os algoritmos de aprendizado de máquina superam os modelos de regressão linear de referência.

Key Words:

Tail Risk, Value at Risk, Sustainability, ESG, Machine Learning

1. Introduction

Environmental, Social and Governance (ESG) and Artificial Intelligence (AI) have been buzzwords dominating the mainstream media and academic literature in recent times. News about climate change, social justice like the gender pay gap and governance issues as for example the discussion of shareholder vs. stakeholder governance have gained traction. Large corporations implement principles supporting conscious and sustainable ESG practices into their corporate identity and governments started passing regulations giving the environment and stakeholders more attention. Such developments have shaped the global markets so strongly that investors started incorporating ESG practices into their strategies.

Recent literature has examined the relationship between ESG disclosure quality and the increase of investments in to ESG finding moderate effects of the disclosure quality on Tobin's Q, ROA and the reduction of downside risk. Effects were larger on extending the ESG investment scale, pushing ESG investment growth especially in firms situated in mandatory corporate social responsibility (CSR) reporting countries (Wen et al., 2022). In the scope of ESG investments in China, studies have found ESG investments to decrease Green House Gas (GHG) emissions in the short and long run decreasing Chinas GHG emissions by 1,391% for a 1% increase in ESG investing (Wang et al., 2023). When looking at relative market value loss, high ESG ranked funds are found to have more significant market value loss compared to low ESG ranked funds (Cerqueti et al., 2021). Critical literature has linked corporate social and financial performance based on ESG scores and found that abnormal returns should no longer be expected from trading portfolios including both high and low rated firms (Halbritter et al., 2015).

The link between investments and technological advancements has been strong since the .com bubble. With the recent developments in the technology sector, led by the publication of large language models through Open AI's Chat APT, AI has secured its long term focus of investors.

Recent literature examining the usage and economic impact of AI propose a significant increase of AI investments using a self-developed measure. It shows firms to have increasing sales, employment and market valuations when investing in AI. Growth predominantly comes from product innovation. Studies suggest AI to have a significant share in firm level development through product innovation (Babina at al., 2024). From a risk management perspective autoregressive conditional Frechet models (AcF) have been used to build a tail risk measurement index. This was used to construct tail risk early warning signals based on machine

learning algorithms. The machine learning based approach overcame traditional risk measures and significantly increased efficiencies in capital markets (Zang and Chen, 2022). Additionally conditional quantile models are in place to capture deep tail risk via memories of sequential data. Sequential neural networks, a form of machine learning model, are applied resulting in capturing the serial dependence structure in conditional volatility and the deep tails of historical prices outperforming the predictive power of traditional models as for example such of the GARCH family (Wu and Xing, 2019).

When combining the topics of machine learning and ESG, research has been conducted on predicting ESG scores using machine learning algorithms (D'Amato et al., 2021; D'Amato et al., 2024). ESG scores were also incorporated when evaluating financial performance (Sharma et al., 2024; D'Amato et al., 2024; Gupta et al., 2021). Additionally extensive research has been conducted on the relationship between ESG and risk management as explained in the literature below. Yet literature is lacking when combining the areas of ESG, risk management and machine learning. Considering this lack of research combined with the importance of ESG, accurate risk modelling and AI (machine learning) this study focuses on the predictability of left-tail risk, measured in Value at Risk (VaR) at the 1% and 5% threshold by ESG scores using machine learning algorithms.

The research of this study is guided by the question: Can ESG scores accurately predict Left-Tail Risk using machine learning algorithms?

To break the research question down further the following hypothesis are examined:

H1: ESG Scores can accurately predict VaR at the 1% level using machine learning algorithms.

H2: The Environmental Pillar Score is most crucial when predicting VaR at the 1% level using machine learning algorithms.

H3: ESG Scores can accurately predict VaR at the 5% level using machine learning algorithms.

H4: The Environmental Pillar Score is most crucial when predicting VaR at the 5% level using machine learning algorithms.

Ultimately this study aims to assess a combination of topics that are both of high priority and importance in the economy and academia and have a profound impact on the way ESG and risk management is valued from the perspective of the investor.

To achieve this goal, this study proceeds with analysing important literature in the scope of risk and returns related to ESG factors. Following a study is conducted approaching to use ESG Scores to predict Left-Tail Risk using various machine learning algorithms. The data of all companies listed in the S&P500 is used and ESG scores of various forms as well as return indexes and risk matrixes such as VaR are examined. The data selection will be described and justified in detail. Next the methodology is presentenced, starting with data cleaning and manipulation, descriptive statistical and visual analysis, and ending with the application six machine learning algorithms onto the data. The study will then present the found results, dive into a discussion and finish off with concluding and taking on limitations of the study.

2. Literature Review

2.1. Importance of ESG Factors

Environmental, social and governance (ESG) factors have been a prominent topic in financial markets from a returns and risk perspective. The following section will dive into literature introducing such topics, relating it to firm valuation, stock returns, executive behaviour, carbon emissions, climate change, direct implications on financial risks, the bond market as well as give an overview of additional and potential future research. Having an overview of the current academic literature on the such subjects will lay the basis for the following study of attempting to predict left-tail risk by using ESG factors in machine learning models.

The fight against climate change has been implemented in societies and businesses for some time. One of the first tangible commitments to stop global warming on a global scale was the 2015 Paris COP 21 climate agreement. Almost 200 parties signed their commitment to keep global warming within 2 °C of the current level. This implicated calling non-governmental firms to action by monitoring carbon dioxide equivalent emissions and implementing ESG strategies (Bolton and Kacperczyk, 2021).

As financial markets are tightly interlinked with firms through investments, the financial industry realized the importance of ESG factors. From a study asking institutional investors in which form and why climate related risks are implemented in investment decisions, the majority states climate risks to have financial implications to their portfolio returns. Over 50% states climate risks to already be material and therefore not to merely be categorized as long term consequences, as prominently believed. Hereby consequences related to regulations are in focus according to the participants. In line with the assumptions from the Paris agreement, investors believe temperatures to rise significantly, having implications on portfolio companies. From a financial risk perspective this should be of great concern for the financial industry, underlining the significance of examining climate risks in financial markets (Krueger et al., 2020). Due to such findings this study will incorporate financial and climate risk by using ESG scores to attempt to predict left tail risk.

2.2. ESG Factors, Firm Valuation and Stock Returns

When relating ESG factors to firm valuation, studies find positive links between ESG disclosures, improved ESG practices and firm performance. This is coherent with finding a positive and statistically significant relationship between corporate social performance (CSP)

and financial performance on firm level. One does have to note that the economic link between CSP and a firm's financial performance was humble (Brooks and Oikonomou, 2018).

By considering industry specific investments Brooks and Oikonomou (2018) find that sin industries, industries emitting significantly more carbon dioxide compared to others, locally are weaker in ESG performance. This implies higher ESG and climate risks, yet excluding such from investment portfolios diminished returns. The market seems to reward the additional risk with higher returns. This is in line with suggesting that effects on financial performance by CSP is misbalanced. Financial risk or a negative impact on financial performance is higher than positive effects on financial performance relative to CSP performance (Brooks and Oikonomou, 2018).

Continuing to look at the social part of ESG factors, research shows that social risk can negatively influence the excess returns for companies listed in the S&P500 and Nasdaq100. When aiming to maximize firm value social issues must therefore be taken into account and firms should aim to lessen such (Cohen, 2023).

Apart from social aspects, studies have shown all ESG factors to be important and impactful for stock performance (Maiti, 2021). All such factors can be found in ESG ratings, which in turn are positively correlated with higher valuations. This is achieved through lower systematic risk and lower cost of capital. Unfortunately this raises the questions of causality. One might argue that higher valued companies have more funds to invest in activities driving ESG scores and not reverse (Giese et al., 2019).

Sticking with ESG ratings but coming back to the priorly mentioned misbalance of risk and returns related to ESG factors, Lööf et al. (2022) show that companies with higher ESG ratings do have less downside risk. On the flipside such companies also show less potential for upside gains (Lööf et al., 2022). Yet when using ESG ratings to identify risk, an improved risk-return profile can be achieved. By excluding low rated ESG stocks as a sign of tail risk, higher portfolio returns can be achieved on average and maximum returns can find new highs. This comes to the clear advantage of the investors as more value is generated (Indrani and Clayman, 2015). As the effect of ESG related risk is discussed in literature applied to the concept of risk and return profiles of portfolio this study will contribute to the literature by assessing the accuracy of predictability to potentially gain new insights for investors.

2.3. Executive Behaviour towards ESG Risk

Looking at executives' or investors' behaviour towards ESG risk and incorporating such into investment decisions, a study by Krueger et al. (2020) finds three motives. Hereby the researchers point out that investors are keen to protect their reputation in the market, respect and incorporate their moral and ethical considerations and meet their legal duties. It is noticeable that all three motives may have financial implications and consequences, yet the first two can move into non-financial areas (Krueger et al., 2020). The loss of reputation is in line with Cohan's (2023) findings of excess returns being negatively impacted by social risk. On this basis, their research suggests providing and improving educations surrounding ESG related topics to investors and stakeholders of the financial industry. Executives should be aware of the consequences of environmental decay on the risk and returns profile of their portfolios (Cohen, 2023). This again is reflected in ESG scores as Indrani's and Clayman's (2025) find higher return companies having higher ESG ratings showing a compelling ex post correlation between ESG Ratings and stock returns (Indrani and Clayman, 2015).

Relating the upside of environmental education and consideration when making investment decisions to a risk perspective, Indrani and Clayman (2025) show that when excluding lower tails ESG stocks from ones portfolio managers can actively decrease risk and improve returns (Indrani and Clayman, 2015). On this note another study finds a significant reduction in the frequency of environmental risk incidents when firms engage into downside risk reductions (Hoepner et al., 2023). This underlines prior statements about the increase in returns and value of firms and the decrease in risk for investors. Nerveless, investing with ESG consciousness seems to be tailored to the risk-averse investor as returns may be capped but downside and tail risk is improved upon (Lööf et al., 2022). Such can use strong environmental performance to decrease the likelihood of adverse events from happening, as it lowers systematic risk of firms (Bax et al., 2023).

When zooming into the governance part of ESG, Gromley et al. (2023) discover that risk taking is highly dependent on executives' compensation. Generally the board of directors tends to decrease managers exposure to stock price volatility as left-tail risk grows. Additionally convexity plays a key role in risk taking. Payoffs with lesser convexity tend to lead to risk diminishing activities, such as diversification, reserving cash, decreasing leverage and R&D spendings (Gormley et al., 2013). Higher convexity in turn leads managers and CEOs to greater risk taking (Gormley et al., 2013). This underlines the importance of considering governance factors when investing and especially when quantifying the risk profile of stocks.

This ties in with Capelle-Blancard's et al. (2019) findings of governance factors potentially having a more noticeable financial impact on stock and firm performance compared to social and environmental issues. Generally it has been observed that engaging shareholders with ESG related issues reduces downside risk measured in value at risk and lower partial moment (Hoepner et al., 2023). Creating awareness and educating executives on the predictability of left tail risk through ESG scores offers the opportunity to influence executives attitude and behaviour towards ESG risk, potentially leading to more informed and favourable decision making.

2.4. Link: Carbon, Climate and Financial Risk

Carbon dioxide emissions and its equivalent play a central role in measuring ESG performance of firms. Bolton and Kacperczyk (2021) argue that investors might expect compensation for investing in and holding stocks of companies with disproportionately high carbon dioxide emissions. Such is rooted in the greater exposure of such firms to carbon risk reflecting negatively on stock returns. In terms of traditional risk factors such firms face larger technological risks as more innovative competitors can gain an advantage when not necessarily relying on fossil fuels but green energy in turn. The literature refers to this as the carbon risk premium hypothesis. When relating this concept to the assumptions of information disbalance in the market, Bolton and Kacperczyk argue that assets are mispriced due to an understatement of risk (Bolton and Kacperczyk, 2021).

Empirically Bolton and Kacperczyk (2021) find that carbon emissions significantly affect stock returns as investors demand a carbon premium. Economically this reflects an annualized increase on returns of 1,8% to 3,1% resulting from a one standard deviation increase in the level of scope 1 emissions. For scope 2 and 3 emissions increases in returns can be higher, as emissions scale greater. Such results were found while controlling for size, book-to-market, momentum as well as further firm characteristics. Hereby the carbon premium is demanded for the total level of emissions and the change in emissions on a year by year basis. Yet there was no correlation found between carbon premiums and the intensity of emissions, meaning the emissions per unit of sales (Bolton and Kacperczyk, 2021). This shows the markets responsiveness to the pure output of emissions independent from company performance.

Carbon premiums and climate uncertainty are also found in the option market. Ilhan et al. (2021) found increased prices of options protecting against downside tail risk for companies with greater carbon emissions. Additionally returns volatility of such firms is larger (Ilhan et

al., 2021). In combination with the findings on stock returns these findings show significant risks and costs arising from having more than average carbon dioxide emissions.

As carbon emissions are a key factor to ESG scores and therefore risk premiums asked in the market, it is crucial to understand their behaviour to left tail risk in various institutional and computational settings, marking the contribution of this study to the literature.

2.5. ESG, Volatility and Tail Risk

In the current literature concerning financial risk management and analysis, tail risk and volatility play a major role. Hereby tail risk or tail dependence is defined as “the amount of dependence in the lower-left-quadrant tail or upper-right-quadrant tail of a bivariate distribution” (Frahm et al., 2005). Practically this refers to likelihood of extreme negative or positive events to happen to a company and therefore influencing stock returns. Technically tails arise when the probability of such events outgrows three standard deviations from the mean of the normal distribution (Bax et al., 2023). Extreme events or shocks can result from product innovations, shifts in international trade and government regulations (Gormley et al., 2013).

Literature is unclear about right or left tails being heavier. Jondeau and Rockinger (2003) find neither tails to be heavier and connect the perception of heavier left tails to the general infrequency of extreme events to occur (Jondeau and Rockinger, 2003). Bax et al. (2023) argues against such statements, claiming left-tails to actually be heavier. More importantly, left tail dependence can serve as a proxy for systemic risk and has shown correlations to corporate investment and risk taking (Bax et al., 2023). Quantifying such shows to have diverse methods. Traditionally percentiles of daily returns within a month are taken but studies have calculated tail risk by averaging daily cross-sectional tail risk improving out of sample predictability therefore paving the way for increase investment success (Faias, 2023).

Drawing the connection to ESG, Shafer and Szado (2020) show that practices in all ESG pillars are crucial for reducing perceived tail risk. Environmental factors play the most crucial role though (Shafer and Szado, 2020). Other literature shows increased left tails risk for superior ESG rated companies (Bax et al., 2023).

In terms of volatility, ESG factors are argued to present lower volatility resulting on decreased risk and higher risk-adjusted returns. Such is argued to be of great relevance for companies and investors due to increased levels of volatility in current markets (Ashwin Kumar et al., 2016). Indrani and Clayman (2015) take this a step further by indicating a significantly strong negative

correlation between ESG ratings and stock volatility. Following their argumentation ESG ratings show the potential to forecast stock risk by reducing average stock specific risk through diversification. Additionally when dropping lower-tail ESG companies from investment portfolios, overall volatility seems to decrease. This is in line with higher risk adjusted return stocks having higher ESG ratings compared to lower risk adjusted return stocks (Indrani and Clayman, 2015).

A

similar concept was applied by Shafer and Szado (2020), who use implied volatility smirk taken from option markets to forecast tail risk. Their research show ESG practices decreasing investors anticipation of tail events and creating a tool for risk protection. Robustness rests came back positive when adjusting for endogeneity of ESG practices and expected tail risk (Shafer and Szado, 2020).

Looking at the bond market, Capelle-Blancard et al. (2019) find a strong negative correlation between ESG performance of firms and sovereign bond spreads. This points to decreased levels of default probabilities as seen in lower bond spreads for top performing companies (Capelle-Blancard et al., 2019).

Adding to the existing literature this research will shine new light of the relationship of ESG scores and left tails risk by using machine learning algorithms as compared to traditional regressions for the prediction of values.

3. Data

In order to approach to answer the research question of ESG scores' predictability of tail risk, relevant data was extracted from the Refinitiv Eikon Excel Plug in. The choice of variables was dictated by the literature explained above. This lead to using Refinitiv Eikons' Environmental Pillar Score, Governance Pillar Score, Social Pillar Score, ESG Score and Total Return Index. Hereby the scores are planned to be used as X variables and Total Return Index as basis for calculation Value at Risk at the 1% level and 5% level ultimately serving as Y variables.

The Environmental Pillar Score abbreviated as ENSCORE in Refinitiv Eikon describes the weighted average relative rating of a company. Such is based on reported environmental information and results from combining three environmental category scores, which consist of underlying key performance indicators. The three environmental category scores are described as Resource Use, Emissions and Innovation. Resource use and innovation carries 29% weighting individually where as emissions are weighted at 41% summing to 100% and therefore making up the Environmental Pillar Score. It is to mention that the emission score is weighted more significantly, potentially casing a bias in the data (Refinitiv, 2022).

The Governance Pillar Score or short CGSCORE symbolizes the weighted average relative governance rating of a company. This score consists of the sub-scores of Management, Shareholders and CSR Strategy. Hereby the weighting varies more significantly as compared to the Environmental Pillar Score. The Management score leads with 63% of the weight, followed by the Shareholder and CSR Strategy score carrying 21% and 16% of the weight respectively(Refinitiv, 2022).

The SOSCORE or Social Pillar score represents the weighted average relative social rating of a firm and consists of four sub-scores. Such are split into Workforce, Human Rights, Community and Product Responsibility and carry the weights of 48%, 13%, 23% and 16% respectively(Refinitiv, 2022).

Cumulatively the Environmental, Governance and Social Pillar Score build the ESG Score. Similarly to the sub-scores of the individual scores, the individual scores are weighted based their level of significance. Hereby the Environmental Pillar Score is in lead with carrying 37% of the weight, followed by the Social Pillar Score of 33% and finally the Governance Pillar Score with 30% (Refinitiv, 2022).

Overall more than 630 ESG measures are calculated on company level. From such 186 are chosen based on their relevancy in measuring and displaying ESG performance in a comprehensible and comparable fashion. The overall scoring and assessment process is characterized by materiality, data availability and relevance of the individual industries (Refinitiv, 2022).

Moving to the base of the Y variables of this study: Total Return Index. Such is abbreviated as RI and shows the theoretical growth.

A return index (RI) is available for individual equities and unit trusts. This shows a theoretical growth in value of a share holding over a specified period, assuming that dividends are re-invested to purchase additional units of an equity or unit trust at the closing price applicable on the ex-dividend date (Refinitiv, 2024).

From 1988 onwards (and from 1973 for US and Canadian stocks), the availability of detailed dividend payment data enables a more realistic method to be used in which the discrete quantity of dividend paid is added to the price on the ex-date of the payment. This will be of focus for this study as the earliest datapoints are generated in the year 2000 (Refinitiv, 2024).

Incorporating such techniques, the following formula was applied:

$$RI_t = RI_{t-1} \times \frac{P_t}{P_{t-1}}$$

Equation 1: Return Index excluding dividends

Except when $t = \text{ex-date of the dividend payment } D_t$ then:

$$RI_t = RI_{t-1} \times \frac{P_t + D_t}{P_{t-1}}$$

Equation 2: Return Index including dividends

Where:

P_t = Price on ex-ante

P_{t-1} = Price on previous day

D_t = Dividend payment associated with ex-date t

Gross dividends are used where available and the calculation ignores tax and re-investment charges. Adjusted closing prices are used throughout to determine price index and hence return index.

The dataset's time frame was chosen from 2000 being the earliest data points to 2024 being the latest. To start with the year 2000 was dictated by the first ESG scores being documented in this year. An effort was made to extend the time frame as long as possible to get the maximum amount of data. In line with this, the data lasts to the most recent year as of 2024.

Due to its availability of ESG scores yearly data was used. For the RI index daily data was used to later calculate the annual VaR at the 1% and 5% level. This step was crucial since the VaR values were calculated using the historical method of taking the values of the distribution at the first and fifth percentile. Doing this enabled to allocate a value for VaR to each year matching the frequency of the ESG scores. For reasons of comparison the returns were annualized and stated on a yearly level by taking the average of daily returns of each year and multiplying such by 252.

In this frequency data for the individual scores and the RI Index were pulled for 500 companies being in the S&P 500 as of 2024. The S&P 500 was chosen due to the maturity, sectorial diversity and data availability of its firms. The dataset proved to incorporate companies from various sectors and have little missing data points in the returns, specifically chosen to potentially increase the quality of the study and avoid dependencies to for example certain sectors.

4. Methodology

As described in the data section a dataset of 500 companies with their respective ESG related scores and RI index were extracted from Refinitiv Eikon for the time frame from the year 2000 until 2024. Following the extraction the data was explored, cleaned, feature engineered and prepared to run various machine learning algorithms with the aim of exploring the accuracy of ESG scores in forecasting tail risk and variance.

Firstly the data was loaded from Excel into Python. Python was chosen for the data analysis due to superior ability and performance in applying machine learning algorithms compared to other computer programs.

The two datasets, one for the ESG score and one for the returns were treated separately at first. The returns dataset was transformed to a data frame, cleaned by renaming columns and adjusting indexes. As the returns data was on daily basis and the VaR at 1% and 5% were to be calculated, two new data frames for each confidence level were created. The values were calculated by taking the first and fifth percentile of the returns of each year separately. This led to having a data frame with yearly values for the confidence levels of all firms. The values were annualized by multiplying with the square root of 252 accordingly.

The same formatting style was applied to the returns data. A new data frame was created and filled with the mean daily returns. Such values were then multiplied by 252 to annualize them. Eventually the data frames for VaR and returns were concatenated and manipulated in the following way.

The data set was formatted to a data frame for ease of use and performance for the algorithms. The data was structured to have the years as columns and the companies with the four data points respectively as the columns. This format was not ideal for the analysis as it is advisable to use unique categories in columns, which was not the case as each company had the same columns repetitively among all the 500. Additionally the columns for the years was mislabelled.

On this basis the column for years was appropriately labelled and the data was reshaped. It was organized to group columns by company and the years, meaning that a block of rows is reserved for each company. The length of the block of rows is dictated by the amount of years data is available for the company. This setup enabled each column to be unique as only one column for the ESG score, the three individual scores, the RI index and two columns for VaR were needed. As a whole this caused for an organized and processual optimized set up for

running machine learning algorithms later on and called for a total of 12200 rows and 7 columns.

After further exploration of the now reshaped datasets, non-numerical or zero values were detected. Such can not be used by the algorithms such were decided to the dropped. This dropped the row count to 8276 as data for the scores was not present for all years on all companies. It was confirmed that the full range of data was consisting of non-null values.

In the following step the data was explored further. Basic statistics for each of the columns, being the scores, RI index and VaR was calculated. The focus was put on receiving further insights from the values mean; median; mode; standard deviation; min; max; 25th, 50th and 75th percentile; count; skewness; and kurtosis.

In the following step the distributions of all variables was analysed graphically. Histograms with a bin size of 50 and density plots were used accordingly. Both types of graphs where applied to all categories of data and visually placed next to each other, aiming to analyse the distribution of the dataset visually and compare trends among the data. Also skewness, kurtosis and outliers were aimed to be detected. To further evaluate the data, especially concerning outliers, boxplots were created.

Following the graphical analysis of the distribution of the data, a correlation matrix was calculated and presented visually using a cool warm heat map. This was done to quantify to priorly visually observed relationships between the data points. Additionally a correlation matrix gave to opportunity to rank and select features as well as test for multicollinearity within the X variables. The representation of the heat map made it easier to detect patters among the data.

Due to the newly gained insights from the correlation matrix, the data again was analysed graphically. This time a pair plot was established to create a matrix structure illustrating each correlation and distribution of the variables in relativity. In other words a grid was established giving each variable a space on the X and Y axis. For each pair of variables an individual coordinate system was created, the data was plotted and regression lines were drawn. This enabled a clear visualisation of relationships, leading to being able to compare the insights gained from the correlation matrix with the visual representation of the data.

From the gained insights the analysis was focussed on the relationship between VaR as the Y variable and the ESG related scores and the X variables. Here regression plots were

created and visualized in a grid of subplots with individual regression lines. This enables the clear visualization and direct comparison of X variables relative to Y and their relative relationship amongst each other.

The latest step was aimed to round of the descriptive analysis by first having looked at descriptive statistics, then proceeding to describing and visualizing the distribution of individual variables, analysing correlations numerically using multiple correlation matrixes and heat maps to finally zoom into the most important relationships graphically, including regression elements.

After finishing the descriptive analysis the research proceeded to applying machine learning algorithms to attempt to predict the Y variable based on the X variables. In first instance the appropriate libraries were downloaded, which were retrieved from sklearn. For the tools the following were used: Standard Scalar, Train Test Split, Mean Squared Error, R² Scores. For the models the following were used: Linear Regression, Ridge, Lasso, Support Vector Machines, Random Forrest Regressors, Gradient Boosting Regressors.

Firstly the data was split into a training and testing set with a distribution of 80% of the data going into the training set and 20% into the testing set. This was performed as a routine task when preparing data for machine learning models. The training set later is aimed to train the individual models on the data whereas the testing set is used to as unseen data to evaluate the performance of the model relative to its input valuables in unknown territories. This finally determines the accuracy of the given model and therefore its predictive power by avoiding biases caused for example from overfitting.

After having split the datasets, each, the training and the testing set was scaled using the standard scalar individually. This step is aimed at bring the data into the same scale meaning that the range of the data is standardized or normalized. This aims at increasing the performance of the models by taking out biases in the data.

At this point the data was ready to be trained onto the models. For all models the same process was repeated. First the models were assigned to a variable in the code. This is not one of the variables in the data but an individual variable in the code. Following the train and test data was fit to the appropriate model using the mentioned variable. Finally a prediction was accomplished by applying the model to the testing data. The accuracy of the models was tested using mean squared errors of and R² variables.

The models were chosen for the following reasons. Since continuous variables were to be predicted and all labels of the data including the predictor variable were known, supervised regression models were used. To create diversity in terms of methodology in order to avoid biases and hyper focusing on similar models a variety of models was applied. Firstly the study deployed a linear regression model to set the baseline for performance and this model is one of the simplest, assuming a linear relationship between the variables.

Moving to slightly more sophisticated models Ridge and Lasso were applied. Both models apply regularization methods, meaning a penalty term is added to the cost function. This enables the model to automate feature selection through the weights in the regression avoiding or at least managing multicollinearity. In the case of ridge the weights of features can be decreased to a minimum if the feature is evaluated not to be crucial. For Lasso the weights can be pushed to zero if the feature is not believed to be important. Both models, in varying intensity, cause the analysis to be focussed on the most important variables, eliminating or minimizing insignificant ones and therefore driving the performance of the model (Owen, 2007).

Support Vector Machines are algorithms usable for regression and classification problems. Appropriate to this research regressions will be of focus. It works by reducing the cost function to find local and global minimums by using a variety of cost functions. The algorithm is capable of applying the quadratic loss function, which corresponds to the ordinary least squares method. The Laplacian loss function can be applied when needing to be sensitive to outliers. When not being aware of the underlying distribution of the data the Huber loss function can be used. An issue proposed by Vapnik with the stated loss function is sparseness. To improve upon this, he proposed the ϵ – insensitive loss function, which out of the four has proven to be the most reliable and therefore most widely used (Gunn 1998, pp. 29-30). In the scope of this research its ability to produce robust results, insensitivity to outliers and ability to deal with sparseness supported the choice to use it.

Similarly Random Forest can be used for both regression and classification problems within supervised machine learning. The mechanisms for both approaches are similar yet distinctive. The algorithm works by developing trees based on random vectors. The predictor value takes on a numerical value as expected for regression exercised. The final random forest predictor is chosen by taking the average of all predicted values of all trees. Hereby it is assumed that the number of tree goes to infinity. By applying this process the random forest

decreases the average error of the individual trees (Breiman, 2001). Using this algorithm as driven by the need for diversity in the methodology. Whereby cost functions are not merely decreased to local or global minimums but the average of a array of results is used leading to improved accuracy and diversity through other methods then used before.

The basic idea behind Gradient Boosting Regressors is to maximally correlate different models as base learners with the negative gradient of the loss function. This aims to provide a more exact prognoses of the Y variable while leaving the researcher the freedom of choice in for example the lass function used. The final choice of loss function is driven by trial and error as the algorithm is capable of fitting different models. While keeping respect to the negative gradient, the most suitable cost function for the given data can be found (Natekin and Knoll, 2013). This rounds up the usage of various algorithms in this study by the ability to deploy versions models giving a suitable extension to the linear regression, regularization and randomization based algorithms described above.

Finally the mean squared error and R^2 were used to evaluate the performance of the individual models. The mean square error was chosen due to its simplicity and commonness within the field of regressions. The R^2 was chosen for determining the goodness of fit.

5. Results

5.1. Descriptive Statistics

The results section of this study is structured by describing descriptive statistics of the study first. This will be followed by looking at the distribution of the X variables, the ESG scores in detail. Afterwards the relationship of the X and Y variable will be discussed by depicting a correlation matrix and various graphs showing the variables in relativity. Finally the results of the applied machine learning algorithms are presented.

	Mean	Median	Mode	Std	Min	Max
Environment_Pillar_Score	44,25	47,04	0,00	29,30	0,00	98,55
Governance_Pillar_Score	56,58	58,71	50,00	21,49	0,62	99,45
Social_Pillar_Score	54,23	55,49	57,51	22,46	0,26	98,26
ESG_Score	52,17	53,68	45,46	20,34	0,60	95,16
VaR_1%	4,08E+05	42.547,76	1.721,43	3,62E+06	55,28	1,17E+08
VaR_5%	4,24E+05	44.165,62	683,08	3,74E+06	73,66	1,19E+08
Returns	7,94E+06	845.133,73	268.686,18	6,89E+07	2.454,43	2,02E+09

	25%	50%	75%	Count	Skewness	Kurtosis
Environment_Pillar_Score	18,90	47,04	69,90	8276	-0,1177	-1,2653
Governance_Pillar_Score	40,91	58,71	73,65	8276	-0,3137	-0,7314
Social_Pillar_Score	36,46	55,49	72,26	8276	-0,1391	-0,9058
ESG_Score	36,10	53,68	69,00	8276	-0,1998	-0,9119
VaR_1%	8.996,16	42.547,76	1,71E+05	8276	20,78	484,12
VaR_5%	9.506,20	44.165,62	1,77E+05	8276	20,58	473,45
Returns	191.825,40	845.133,73	3,26E+06	8276	19,83	433,57

Table 1: Descriptive statistics of all variables, cleaned from null variables but not from outliers

In the columns of table one the descriptive statistics for the variables used on the discussed dataset are listed. The first four rows describe the X variables, which are the Environmental Pillar Score, Governance Pillar Score, Social Pillar Score and ESG Score. VaR 1% and 5% stands for the respective confidence levels of annualized VaR and Returns describes the annualized RI index.

All variables have the same count of 8276 values, indicating that no null variables are existing in the dataset. This is an important criterium for running the machine learning algorithms most efficiently with the given data. Proceeding to the mean, the Governance, Social and ESG score are within a range of 52 to 56 and therefore distributed relatively similarly. The Environmental score falls under 50 with a mean of 44 indicating shortcomings of firms within environmental practices. Considering that ESG scores range form 0 to 100 it becomes evident that all categories but the environmental score fall withing in the top 50%. This may have an effect on the final prediction of the Y variables due to the superior weight of the Environmental score within the total score and the theorised importance of the environmental score derived from the relevant literature.

For the Environmental, Governance and ESG scores the median is slightly higher then the mean, indicating skewness to the left side of the distribution. As for the Social Pillar score mean and median are close together indicating more symmetry in the distribution. For the mode of 0 the Environmental score, showing that the most common value is 0, the data indicates either extremely low values in the respective category or missing data, which was filled with zeros from the source. This is to be monitored as this can effect the predictions of the models later on. For the governance and ESG score the mode is lower then the mean and median pointing the distribution to be left skewed. For the Social Pillar score the mode is higher then the mean and median, which are close together, indicating skewness to the right in a slight fashion.

The standard deviation, ranging from 20 to just shy of 30 suggests considerable variability in the Environmental score, having the highest standard deviation out of the four. This can be traced back to the mode of zero, indicating frequent zero values and therefore more variability in the dataset. The remaining scores show less variability with scores ranging from 20 to 22.46. The standard deviation of the ESG score is the lowest, which is in line of its definition as it consist out of a combination of the other scores.

Moving to the ranges of the scores. Such are also in line with their definitions, was explained the data section as they range from zero or close to zero to 99,45 almost hitting the perfect score of 100. In the skewness of all scores the assumptions from analysing the mean, median and mode are reflected. The skewness of all four scores points to the left indicating fatter tails to the left end of the distribution. Hereby the Governance Pillar score takes the lead in having the most significant skewness out of the four, which is in line with its median being higher then its mean.

In terms of kurtosis all values of the X variables are negative indicating the distribution to not be as peaked as seen in normal Gaussian distributions.

Moving to the Y and support variables, of VaR at the 1% and 5% confidence level and the RI index. The mean of all three variables is significantly higher then the median. Fort the confidence levels of VaR both mean and median are within similar ranges. The values of the 1% level are lower compared to the 5% level. The significantly lower median compared to mean indicating the distribution to be skewed to the right with the possibility for heady tails. It also indicates outliers in the dataset as there is the possibility for single extreme values to pull up the mean. Considering the nature of VaR to indicate extreme losses, these results are not atypical. The same is to be seen with the returns. Such findings are in line with each other as VaR was calculated from the returns. This created the reasoning for cutting outliers later in the study as indicated in the methodological part.

The arguments presented for the existence of outliers in the dataset is underlined my the most frequently found values represented by the mode of the variables being significantly lower then both mean and median in all three variables. Similarly as in the ESG scores but much more extreme, the low mode might point to fatter tails to the right side of the distribution. This is in line with the type of data and the indication of outliers as some extreme values can alter the distribution of the data significantly.

The great variation on the distribution of the VaR at 1%, 5% and the returns carries forward by high standard deviations in all three variables. The standard deviation of the confidence levels of VaR again shows a similar range compared to other descriptive measurements of the two variables underlining the consistency of the data in its distribution. For the returns the standard deviation is significantly higher, which is in line with the potential for extreme values evident from the analysis of the mean, median and mode.

Also the ranges of VaR at both confidence levels are similar. Both are extreme. This is not atypical for financial data and especially data looking at extreme events with having found outliers of prior analysis. The same is to be seen for the returns data. The significantly positive skewness of VaR at both confidence levels underlines prior findings. The predominant part of the risk exposure or losses are not extreme yet there is the potential for such, which reflected the nature of extreme event to be able to occur especially when considering that the time frame of the data ranges through multiple financial downturns. The distribution is heavily clustered around the right tail and therefore lower end of the distribution. The same is to be seen for the returns.

The domination of extreme values carries forward into the kurtosis of the three variables. The kurtosis is considerably high in all cases, indicating the distributions to peak with fat tails. This is in line with the data describing financial risk in various times of volatility in the markets. Again the majority of risk is moderate but strong outliers make the tails peak.

To deepen the found results from the descriptive statistics, the following part will describe histograms, density plots and box plots of the distributions before moving into analysing the correlations between the variables.

5.2. Descriptive Chart Analysis

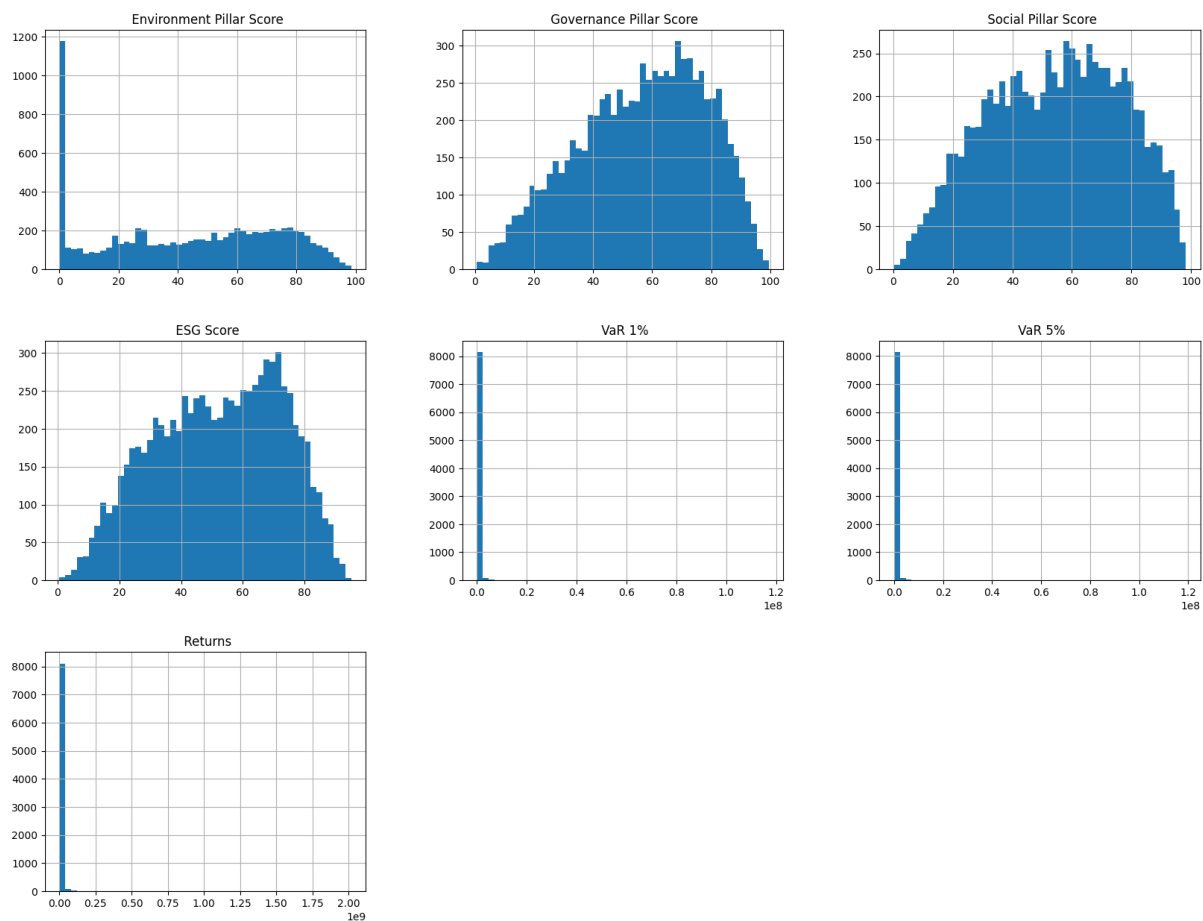


Figure 1: Histograms of the distributions with a bin size of 50

The histograms reflect the findings from the descriptive statistics. For the Environmental Pillar Score the abundance of zero values and therefore its mode of zero is evident by the strong peak to the very left of the chart visualizing zero values. When looking at the remaining distribution such is skewed slightly to the left with more values exceeding the mean. Also the range of scores is suitable as such are distributed between 0 and 100.

The governance pillar score appears to be distributed almost normally with a slight skewness to the left. Slight peaks in the values surrounding 50 and 60 are to be seen, which is in line with its mode of 50 as well as mean and median ranging from 56 to 59. With the governance score having the highest skewness but lowest kurtosis out of the four, therefore indicating to skinnier tails is reflected in the seen distributions.

The social score appears to be distributed similarly yet with higher levels of kurtosis and lower levels of skewness as the tails are fatter but more evenly distributed around the mean. Peaks in density of values around a score of 60 are to be seen, which is to be expected from the

respective mean and median of the distribution. Similar results can be seen in the distribution of the ESG score yet the level of skewness is slightly higher compared to the other scores, leading the distribution further into higher values compared to the mean.

For the VaR and returns data, the strong population along the right tails becomes evident due to great density visually forming around the zero axis. Yet data is varying out to over 1,2 le8 values showing that most values stay within considerable risk but strong breakouts in cost do exist. Such underlines the insights gained from the descriptive statistics where the combination of mean, median, mode, standard deviation and the ranges indicated a similar distribution.

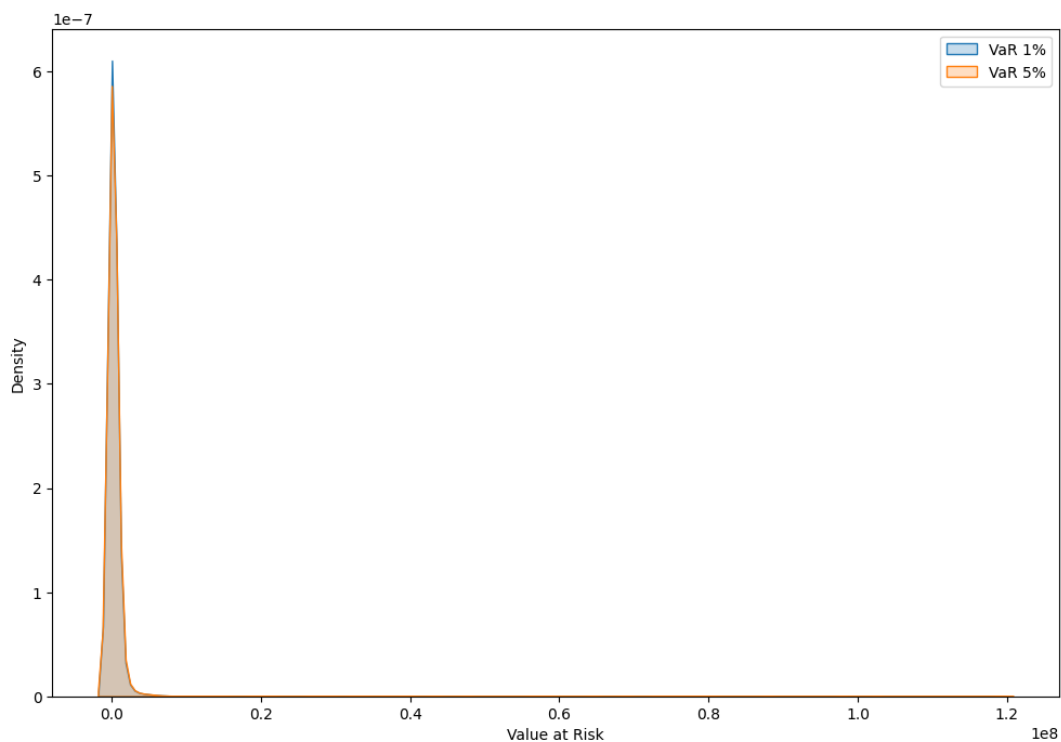


Figure 2: Density Plot of VaR at 1% and VaR at 5% stacked

Deepening this analysis, figure 2 shows the two confidence levels of 1% and 5% in one plot. This shows more clearly that the distribution centres around zero but comes up over 1,2 le8 marking extreme values.

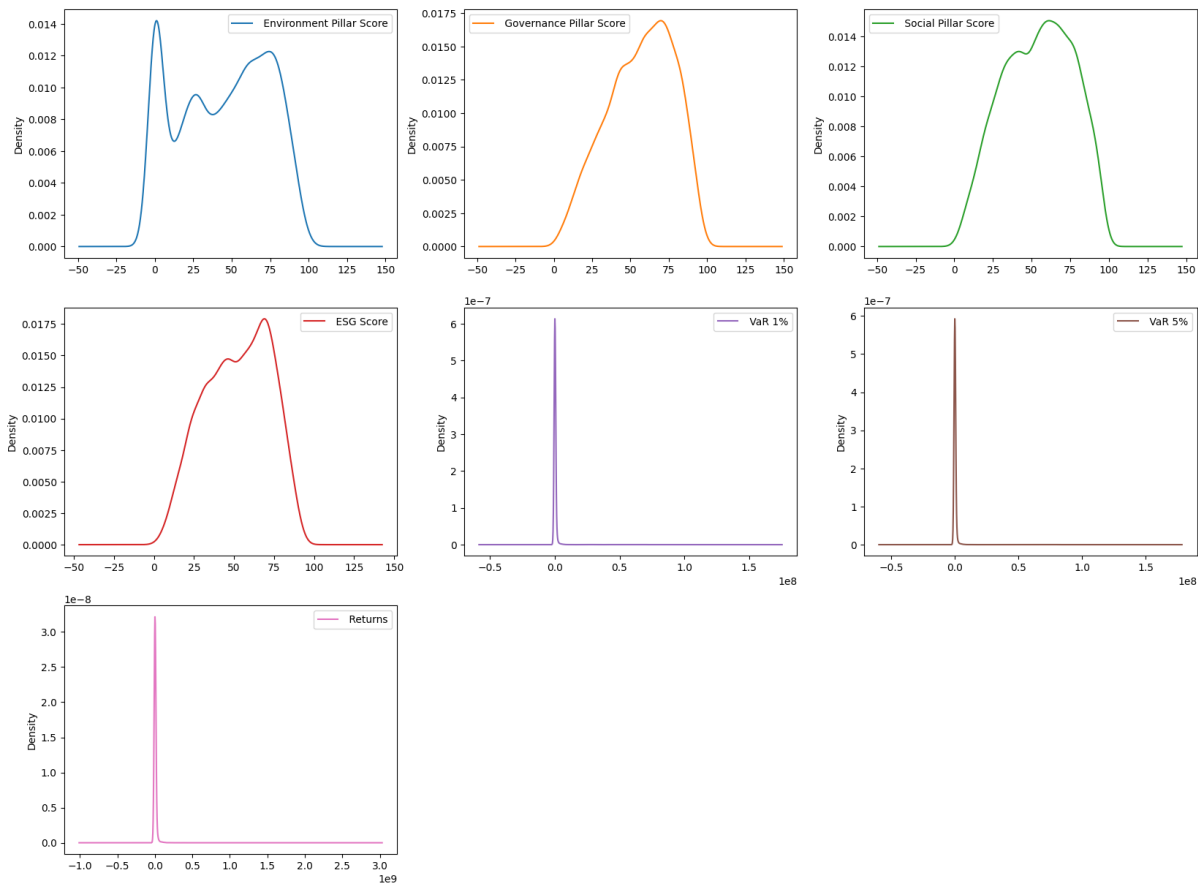


Figure 3: Density plots of the distributions of the variables

Following the style of figure 2, figure 3 shows the strong presence of zero values in the distribution of the Environmental Pillar score. One can observe the peak in values right from the start, which is not evident in the distributions of the other scores. Here more than in the histograms, the similar distribution of the all scores comes into play when ignoring the zero values of the Environmental Pillar score. All scores see their peaks on the left side of the distribution, meaning above the mean underlying the observe skewness. One has to note the different scale of Y axis of the plot of the Environmental and Social score plot compared to the Governance and ESG score plot.

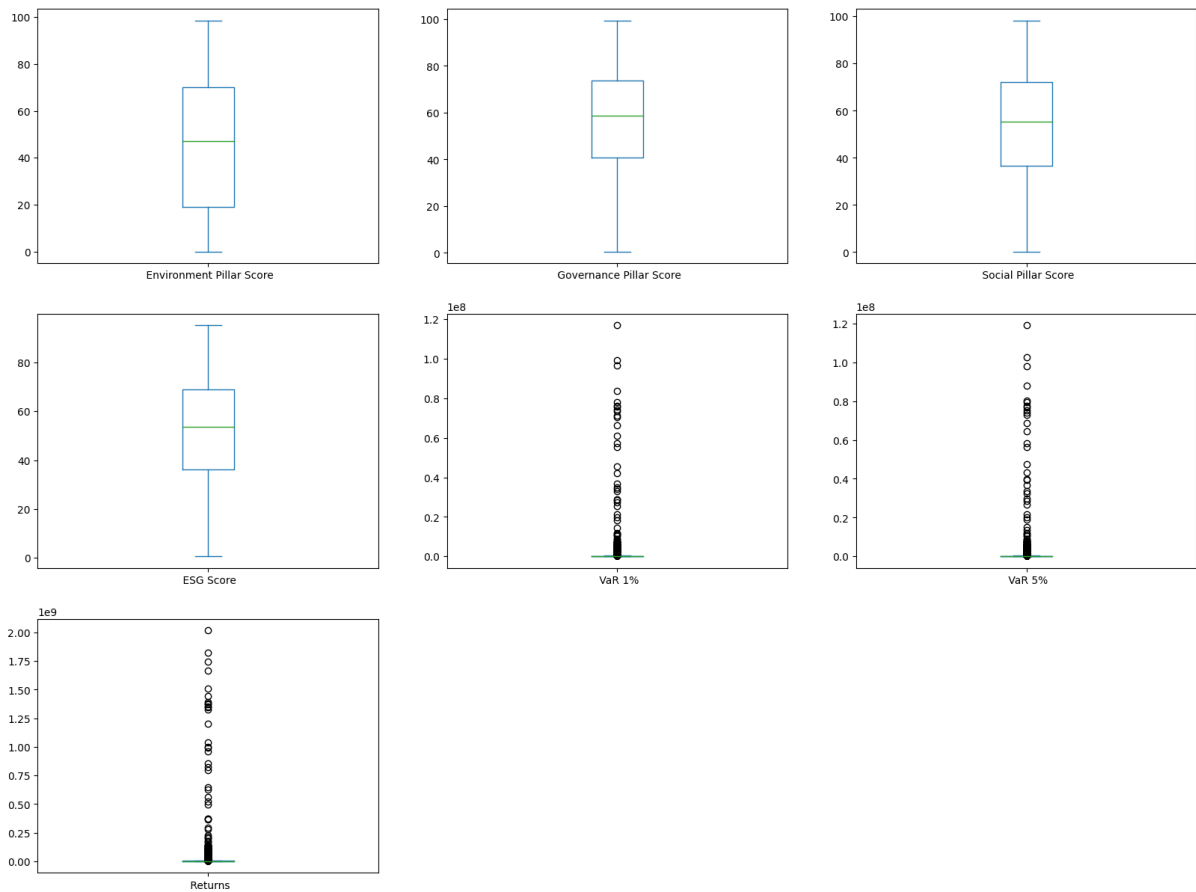


Figure 4: Box plot of the distributions of all observed variables

Visualizing the distributions of the variables through box plots is in line with the prior visualizations by showing similar distributions in the ESG related scores. Medians appear to be balanced across all four cores and the interquartile ranges are symmetrical surrounding the median, here marked as a green line. The lack of outliers makes all scores being in the correct ranges, as mentioned in the range during the discussion of the descriptive statistics. The symmetry and range of whiskers also underlines this.

Moving to the box plots on VaR, the extreme values become evident immediately. The green line, indicating the median, is very low compared to the total distribution. This in combination with the box of the interquartile ranges not being visible shows that the majority of the distribution is within the lower ranges and that 50% of the data is relatively close together. The strong presence of outliers again shows that most losses are relatively low and extreme events do exist. Such range up over 1.2×10^8 as discussed before when looking at the density plot overlapping VaR at the 1% and 5% confidence level. Similar distributions are seen in the returns.

5.3. Correlation Analysis

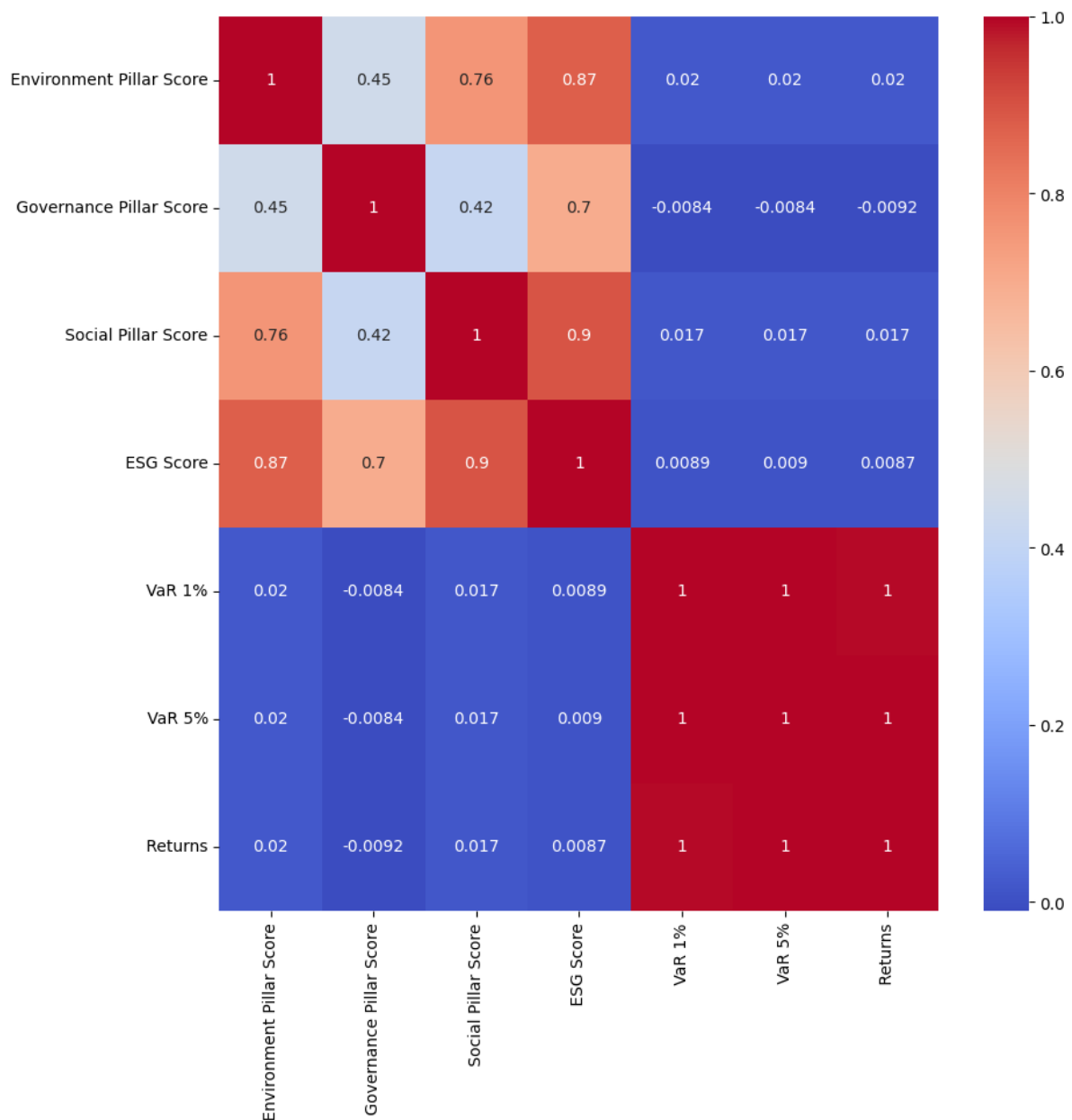


Figure 5: Correlation Matrix of the analysed variables in style of a heat map

Leaving the descriptive chart analysis and focussing on the correlation between features, this section will focus on the correlation between the X variables, being the four ESG related scores and the Y variables being VaR at the 1% and 5% confidence level.

In general correlations between X and Y variables are low. Most of them are positive but the low scores suggest no clear linear relationship between ESG related scores and left-tail risk expressed in VaR.

The Environmental Pillar score as the highest positive correlation to VaR at both confidence levels suggesting that of the four the Environmental scores influence VaR the most, yet correlations are low. The Governance Pillar scores shows the only negative correlation to VaR yet this is also close to zero. The Social Pillar scores has the second highest correlation to VaR, as the ESG score, as expected falls into the middle of the remaining scores. Also from the colour coding of the heatmap, the insignificance of correlation becomes evident, since all values are within the deep blue. Just for reference the returns and VaR stand in perfect correlation to each other, as expected, as VaR is calculated on the base of returns. The correlation between the ESG related scores and returns is similarly low compared to VaR.

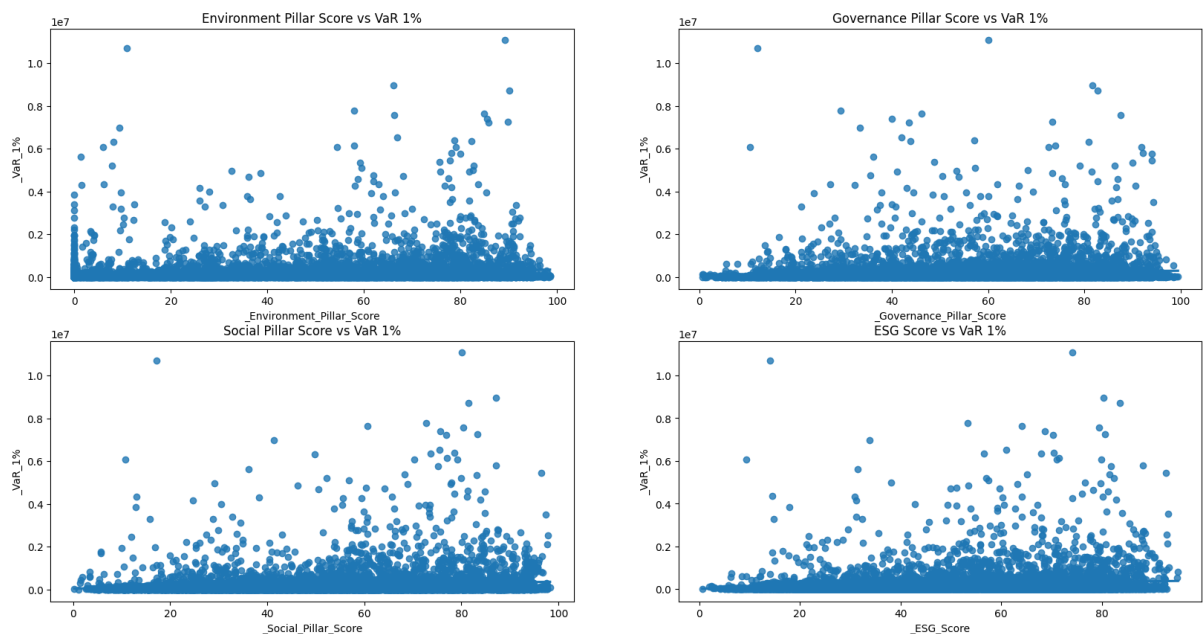


Figure 6: Regression plot of the ESG related scores and VaR at 1% confidence level

When analysing the relationship of the ESG related scores relative to VaR at the 1% confidence level most datapoints scatter along the lower VaR values with exceptions describing more extreme events. Regression lines for all plots are flat to mildly increasing making it difficult to read a clear upward or downward trend in a companies VaR when ESG related scores change for the better or worse. This underlines prior findings in the distribution of VaR but does not underline robust findings in the relationship between ESG related factors and financial risk. The same is to be observed on the relationship of the X variables to VaR at 5% shown in figure 7.

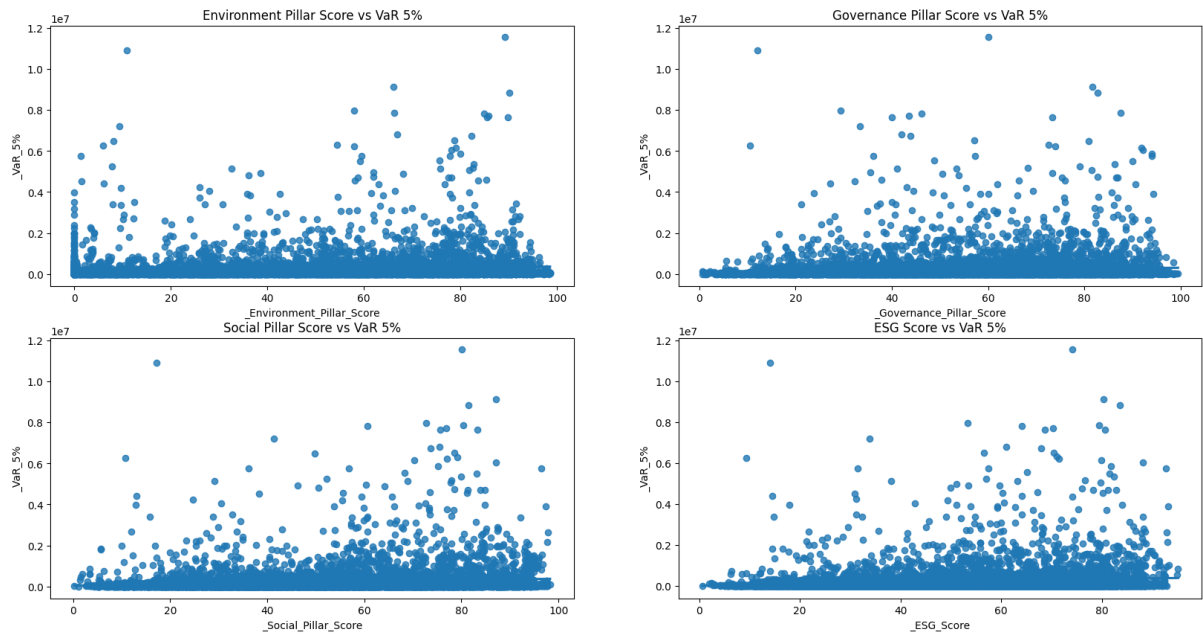


Figure 7: Regression plot of the ESG related scores and VaR at 5% confidence level

5.4. Findings from machine learning based predictions

VaR 1%			VaR 5%		
Model	MSE	R ²	Model	MSE	R ²
Linear Regression	3,34E+11	0,017688	Linear Regression	3,64E+11	0,017915
Ridge Regression	3,34E+11	0,017647	Ridge Regression	3,64E+11	0,017875
Lasso Regression	3,34E+11	0,017687	Lasso Regression	3,64E+11	0,017914
SVR	3,71E+11	-0,08847	SVR	4,03E+11	-0,08828
Random Forest	3,34E+11	0,019414	Random Forest	3,66E+11	0,013115
Gradient Boosting	3,39E+11	0,003773	Gradient Boosting	3,69E+11	0,003805

Table 2: Results from machine learning based predictions on VaR at the 1% and 5% confidence level

As for the results of the machine learning models approaching to predict VaR at both 1% and 5% confidence level general accuracy was low, which is to be derived from the high mean squared errors and low R² values. The linear regression as well as ridge and lasso show similar mean squares errors and R² scores for both confidence levels. Neither the lasso or ridge model outperformed the linear model for both confidence levels, which is to be seen in the higher mean squared error and lower R² score. Whereby the error is of great size and the model is merely able to explain 1,7% to 1,8% of the variance of the Y variables. One can conclude that regularization did not significantly improve the accuracy of the models, which

traced back to the relatively small size of input variables or the ESG related scores acting as a summary of variables.

The support vector machines show the lowest accuracy of all applied models with the highest mean squared error and R^2 on the negative spectrum. This can be traced back to the lack of correlation between the X and Y variables or the inappropriateness of the kernel used with the model. The random forest model shows the best performance out of the six. Mean squared errors are still comparatively high and R^2 is low but the model performs best in explaining the accuracy of VaR at least at the 1% level. Gradient boosting was except for the support vector machines at the 5% level the worst performing model in explaining the error and variance of the models used.

Considering the priorly presented results, non of the null hypothesis can be rejected as the accuracy of the machine learning models used to predict VaR at the 1% and 5% level using ESG related scores was low. Against the results of the presented literature this study found no strong linear correlation between ESG related scores and VaR based on return indexes for S&P500 firms from 2000 to 2024. One can say that the Environmental score had the highest correlation and predictability out of all the scores but it was still too low to be significant. The poor performance of the models can be accredited to the high variability of the VaR data and its challenging predictability due its nature of capturing extreme events.

6. Conclusion

Attempting to predict VaR at the 1% and 5% confidence level using various machine learning algorithms shows that ESG related scores are an inaccurate predictor by itself. None of the scores nor algorithms proved to be accurate to a satisfying level, measured in the high mean squared errors and low R^2 values. One is able to see that the Environmental Pillar score has the highest correlation with VaR out of the scores and that Random Forest Regressors is the most accurate algorithm in predicting VaR using ESG related scores.

This study entails retrieving the Environmental, Governance and Social Pillar score as well as the ESG score and Return Index for S&P500 companies from year of 2000 to 2024 from Refinitiv Eikon. The data was cleaned and analysed in Python including calculating VaR at the 1% and 5% confidence level, cutting null values and outliers and running a detailed analysis on the descriptive statistics and distribution of each variable. Based on the gained insights the ESG related scores were used as X variables to be trained by using six machine learning algorithms. The algorithms were put in place appropriately by being trained on the data and pursued to make predictions based on the pattern detected. Finally the performance of the algorithms using ESG related scores to predict VaR was measured and discussed in context of the prevailing literature.

7. Limitations & Future Outlook

The study was limited by the timeframe of scores available. The earliest of scores come from the year 2000, whereas stocks have been listed on various indexes for significantly longer. Meaning that even though returns can be retrieved for a longer time frames the ESG related scores hindered the model to make more accurate predictions due to the lack of scores prior to 2000. Additionally the data showed a significant amount of zero values cutting the dataset more. For future research it would be beneficial to retrieve more data, looking at additional indexes or smaller companies for a pure data availability perspective.

ESG related scores summarize a variety of data points and are used for their simplicity in terms of data retrieval and maintenance. This caused the analysis and implementation of machine learning models to be efficient. Yet when considering their low predictive power, future studies can analyse the input variables of the scores more gradually and put such into the context of predicting VaR. The same logic goes for choosing VaR as the predictor variable for left-tail risk. Other matrixes such as variance could be taken into consideration on the future.

Publication bibliography

Ashwin Kumar, N. C.; Smith, Camille; Badis, Leïla; Wang, Nan; Ambrosy, Paz; Tavares, Rodrigo (2016): ESG factors and risk-adjusted performance: a new quantitative model. In *Journal of Sustainable Finance & Investment* 6 (4), pp. 292–300. DOI: 10.1080/20430795.2016.1234909.

Bax, Karoline; Sahin, Özge; Czado, Claudia; Paterlini, Sandra (2023): ESG, risk, and (tail) dependence. In *International Review of Financial Analysis* 87, p. 102513. DOI: 10.1016/j.irfa.2023.102513.

Bolton, Patrick; Kacperczyk, Marcin (2021): Do investors care about carbon risk? In *Journal of Financial Economics* 142 (2), pp. 517–549. DOI: 10.1016/j.jfineco.2021.05.008.

Brooks, Chris; Oikonomou, Ioannis (2018): The effects of environmental, social and governance disclosures and performance on firm value: A review of the literature in accounting and finance. In *The British Accounting Review* 50 (1), pp. 1–15. DOI: 10.1016/j.bar.2017.11.005.

Capelle-Blancard, Gunther; Crifo, Patricia; Diaye, Marc-Arthur; Oueghlissi, Rim; Scholtens, Bert (2019): Sovereign bond yield spreads and sustainability: An empirical analysis of OECD countries. In *Journal of Banking & Finance* 98, pp. 156–169. DOI: 10.1016/j.jbankfin.2018.11.011.

Cohen, Gil (2023): The impact of ESG risks on corporate value. In *Rev Quant Finan Acc* 60 (4), pp. 1451–1468. DOI: 10.1007/s11156-023-01135-6.

Faias, José Afonso (2023): Predicting the equity risk premium using the smooth cross-sectional tail risk: The importance of correlation. In *Journal of Financial Markets* 63, pp. 100769–100791. DOI: 10.1016/j.finmar.2022.100769.

Frahm, Gabriel; Junker, Markus; Schmidt, Rafael (2005): Estimating the tail-dependence coefficient: Properties and pitfalls. In *Insurance: Mathematics and Economics* 37 (1), pp. 80–100. DOI: 10.1016/j.insmatheco.2005.05.008.

Giese, Lee, Melas, Nagy, Nishikawa (2019): Foundations of ESG Investing: How ESG Affects Equity Valuation, Risk, and Performance.

Gormley, Todd A.; Matsa, David A.; Milbourn, Todd (2013): CEO compensation and corporate risk: Evidence from a natural experiment. In *Journal of Accounting and Economics* 56 (2-3), pp. 79–101. DOI: 10.1016/j.jacceco.2013.08.001.

Hoepner, Andreas; Oikonomou, Ioannis; Sautner, Zacharias; Starks, Laura T.; Zhou, Xiao Y. (2023): ESG Shareholder Engagement and Downside Risk, pp. 1–51.

Ilhan, Emirhan; Sautner, Zacharias; Vilkov, Grigory (2021): Carbon Tail Risk. In *The Review of Financial Studies* 34 (3), pp. 1540–1571. DOI: 10.1093/rfs/hhaa071.

de Indrani; Clayman, Michelle R. (2015): The Benefits of Socially Responsible Investing: An Active Manager’s Perspective, pp. 49–72.

Jondeau, Eric; Rockinger, Michael (2003): Testing for differences in the tails of stock-market returns. In *Journal of Empirical Finance* 10 (5), pp. 559–581. DOI: 10.1016/S0927-5398(03)00005-7.

Krueger, Philipp; Sautner, Zacharias; Starks, Laura T. (2020): The Importance of Climate Risks for Institutional Investors. In *The Review of Financial Studies* 33 (3), pp. 1067–1111. DOI: 10.1093/rfs/hhz137.

Löf, Hans; Sahamkhadam, Maziar; Stephan, Andreas (2022): Is Corporate Social Responsibility investing a free lunch? The relationship between ESG, tail risk, and upside potential of stocks before and during the COVID-19 crisis. In *Finance research letters* 46, p. 102499. DOI: 10.1016/j.frl.2021.102499.

Maiti, Moinak (2021): Is ESG the succeeding risk factor? In *Journal of Sustainable Finance & Investment* 11 (3), pp. 199–213. DOI: 10.1080/20430795.2020.1723380.

Refinitiv (2024): Refinitiv Eikon - Microsoft Excel. Version 2024. 2024: Refinitiv.

Refinitiv (2022): Environmental, Social and Governance (ESG) Scores from Refinitiv. London.

Shafer, Michael; Szado, Edward (2020): Environmental, social, and governance practices and perceived tail risk. In *Accounting & Finance* 60 (4), pp. 4195–4224. DOI: 10.1111/acfi.12541.

Sharma, Utkarsh; Gupta, Akshat; Gupta, Sandeep Kumar (2024): The pertinence of incorporating ESG ratings to make investment decisions: a quantitative analysis using machine learning. In *Journal of Sustainable Finance & Investment* 14 (1), pp. 184–198. DOI: 10.1080/20430795.2021.2013151.

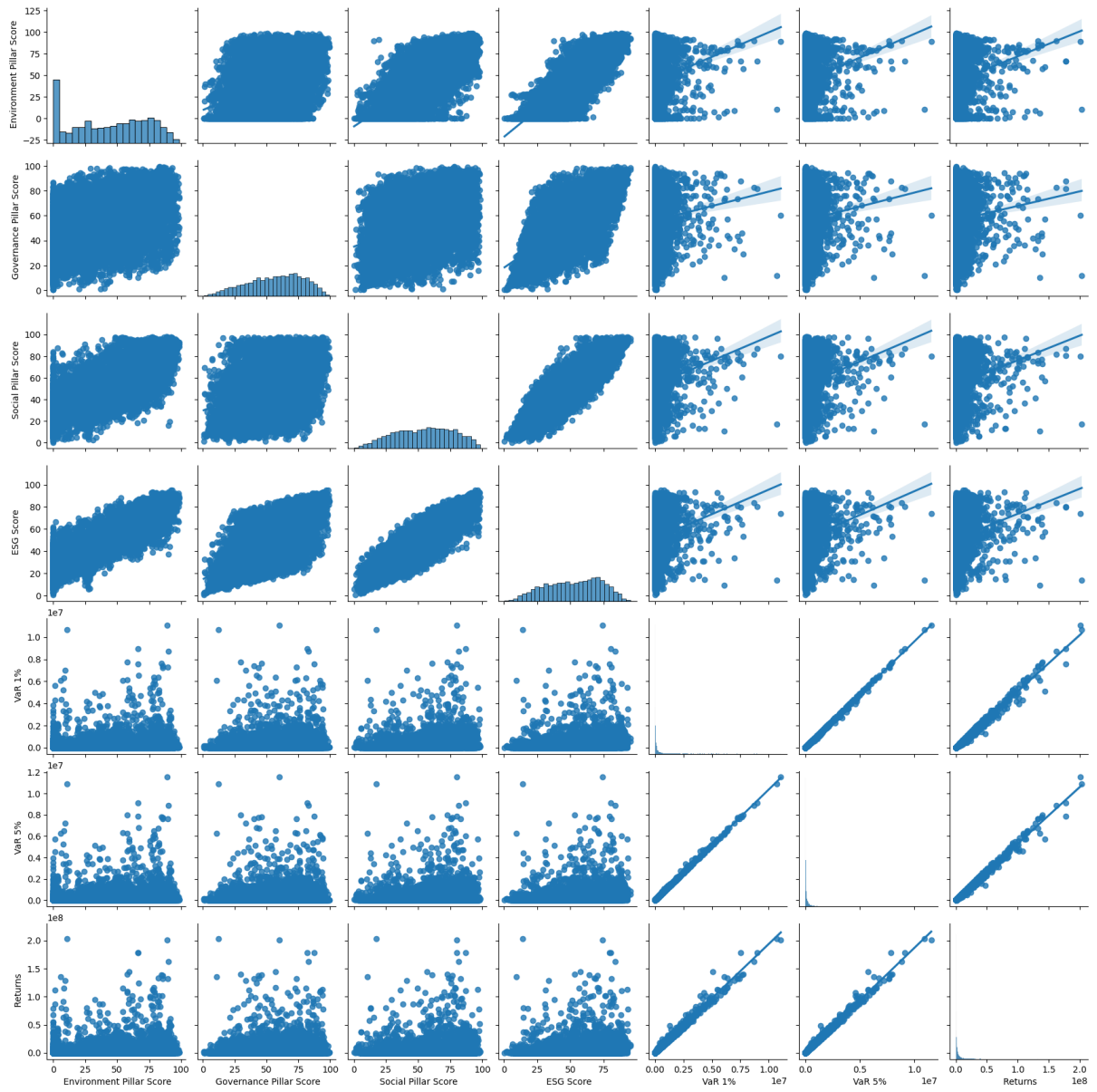
Wang, Xiaoyuan; Wang, Jiahaoran; Guan, Weimin; Taghizadeh-Hesary, Farhad (2023): Role of ESG investments in achieving COP-26 targets. In *Energy Economics* 123, p. 106757. DOI: 10.1016/j.eneco.2023.106757.

Wen, Hui; Ho, Ken C.; Gao, Jijun; Yu, Li (2022): The fundamental effects of ESG disclosure quality in boosting the growth of ESG investing. In *Journal of International Financial Markets, Institutions and Money* 81, p. 101655. DOI: 10.1016/j.intfin.2022.101655.

Wu, Qi; Yan, Xing (2019): Capturing deep tail risk via sequential learning of quantile dynamics. In *Journal of Economic Dynamics and Control* 109, p. 103771. DOI: 10.1016/j.jedc.2019.103771.

Zhang, Zongxin; Chen, Ying (2022): Tail Risk Early Warning System for Capital Markets Based on Machine Learning Algorithms. In *Computational Economics* 60 (3), pp. 901–923. DOI: 10.1007/s10614-021-10171-0.

Appendix



Sworn Declaration

Being aware that making a false declaration is an offence punishable by law and being aware of the legal consequences of this for the examination process, which go as far as exclusion from it in the event of cheating or the attempt to cheat in accordance with the Examination Rules of ESCP Europe Wirtschaftshochschule Berlin e.V. (“ESCP Business School”), I hereby undertake the following sworn declaration to ESCP Business School:

I have written the thesis entitled

“Tail-Risk and Sustainability: Can ESG scores accurately predict Value at Risk? – A machine learning based approach”

entirely by myself and independently and have used no other than the sources and resources named. Further, all the quotations and citations – including tables, maps and figures – are clearly marked as such. This declaration includes the data accessed on the internet. The thesis has not been presented to any other examining body; nor has it been published in whole or in part.

Furthermore, I grant ESCP Business School the right, under the provisions of §31 of German copyright law (Urheberrechtsgesetz), to use the thesis for the purpose of marking and assessing it as well as examining it from the point of view of plagiarism. This right to use the thesis includes the right of ESCP Business School to commission a third party to undertake the plagiarism check on behalf of ESCP Business School. The third party is subject to the same legal requirements on data protection and confidentiality as ESCP Business School.

Mohme, Ulrich

Name, First Name

Berlin, 20.05.2024

Place, Date

Signature

Additional Affirmation

Hereby I confirm that I will not give or sell this thesis for publication to any institution. Furthermore, I will not publish this thesis on the Internet.

In the case that I plan to publish this thesis, I can only do so with written consent of my supervisor.

Mohme, Ulrich

.....

[Name, First Name]

Berlin, 20.05.2024

.....

[Place, Date & Signature]