Vrije Universiteit Amsterdam

Master Thesis

Quantifying the relation between ESG and financial performance using machine learning

Author: Nina Hofstede (2674939)

1st supervisor: Paulo Serra
daily supervisor: Niels Hofman (KPMG, Sofy)
2nd reader: Mark Hoogendoorn

July 30, 2024

Abstract

Over the past few years, ESG (Environmental, Social, and Governance) considerations have become increasingly important in the corporate world. Due to laws and regulations there is a growing driving force for companies to improve their ESG efforts. The European Union's Sustainable Finance Disclosure Regulation (SFDR) and the Corporate Sustainability Reporting Directive (CSRD) are examples of regulations mandating transparency and accountability in ESG performance. Non-compliance with these regulations can lead to penalties, while strong ESG practices can improve a company's financial performance and reputation.

This thesis investigates the relationship between Environmental, Social, and Governance (ESG) performance and financial performance in the corporate world, with a specific focus on machine learning to quantify this relationship. Understanding the effect of ESG performance on financial performance becomes crucial, as ESG becomes more important in business plans and investment decisions. This thesis, conducted in collaboration with KPMG, investigates specific ESG factors and their impact on several financial indicators.

Data for the research was obtained from the LSEG database, with specific emphasis on categories found crucial by KPMG. Various machine learning models, including Linear Regression, Random Forest, Long Short-Term Memory (LSTM), and AutoRegressive Integrated Moving Average (ARIMA), were applied to identify the most important ESG factors and predict financial performance.

The most important features found by Linear Regression and Random Forest were features based on the Human Rights Score, Management Score and CSR Strategy Score. Here Random Forest outperformed Linear Regression in all cases, suggesting a complex, non-linear relationship between ESG performance and financial performance. For the LSTM and ARIMA methods four different scenarios were applied to determine which ESG features to incorporate in the

predictive models. The results indicate that while ESG factors such as Human Rights, Resource Use and Eco-Design may have potential effects, the predictive models incorporating these factors as exogenous variables did not show significant improvement over base models in terms of RMSE.

Contents

List of Figures

List of Tables

1

Introduction

In this introduction the context of the research will be explained together with its relevance to the host company. It will also present the research question and subquestions, followed by an overview of the chapters.

1.1 Context of the Research

Over the past few years ESG (environmental, social, governance) has gained great attention in the corporate world (1). On a daily basis ESG has become crucial in the context of business strategy and investment decisions. Environmentally, there is a strong focus on issues such as climate change and pollution. Socially, companies are increasingly prioritizing their social footprint, including human rights. Lastly, strong governance practices ensure transparency and build trust (2).

Worldwide more and more laws and regulations around ESG are enforced. For instance the Sustainable Finance Disclosure Regulation (SFDR) introduced by the European Union, which *'imposes mandatory ESG disclosure obligations for asset managers and other financial markets participants with substantive provisions of the regulation effective from 10 March 2021'* (3). Other than that there is the Corporate Sustainability Reporting Directive (CSRD). *'In addition to disclosing information on policies and initiatives, the CSRD requires organizations to set targets, select a baseline and report progress towards these targets'* (4). These laws and regulations aim to increase transparency, protect investors, and drive companies towards more sustainable and ethical practices. When not complying to these laws and regulations this can result in penalties.

Not only is it of great importance for companies to comply with laws and regulation to prevent getting penalties, there is an increasing amount of evidence that high quality

ESG performance can lead to financial benefits (1)(5)(6). Companies that actively control their ESG impact often experience improved reputations, operational efficiencies, and increased interest from investors who prioritize sustainability. These advantages may results into better financial performance. The main goal of this research is to investigate the relationship between ESG performance and financial performance. Specifically, it aims to determine which ESG factors influence financial outcomes and how these factors can be used to predict financial performance.

The increased focus on ESG reflects a structural change rather than just a temporary trend. It becomes more and more clear that ESG compliance are crucial for long term success. Stakeholder expectations grow, together with the amount of laws and regulations on ESG (7), which makes it a crucial area to focus on for companies.

1.2 Problem Statement

While there is already significant amount of research on the relationship between ESG and financial performance, these studies tend to adopt a broad approach, primarily focusing on the three overarching pillars of ESG. Such an approach may fail to capture the nuanced impact of specific ESG components on financial performance. This research aims to address this gap by focusing more on detailed and specific categories within the three pillars. The goal is to determine which ESG factors most influence financial performance and use this information to provide insights for ESG solutions and client advice. Other than that the effect of ESG on different financial indicators will be investigated, this to provide a more comprehensive understanding of the relation between ESG and financial performance.

Research Question:

‘What is the relationship between ESG (Environmental, Social, Governance) performance and financial performance of a company, and how can machine learning models quantify this relationship?’

Subquestions:

- Can ESG performance help predict financial performance more accurately?
- What ESG factors contribute to the prediction of financial performance?
- Which financial performance can be predicted based on ESG performance?

- Which selected method(s) works best to quantify the relationship between ESG and financial performance?

1.3 Methodology

This research will use various machine learning methods to investigate the relation between ESG and financial performance. First of all, data is obtained from the LSEG database. The ESG columns obtained from this database are categories within the three ESG pillar, with particular emphasis on categories identified as crucial by KPMG. The data will be cleaned, and feature engineering will be applied.

The next step is to apply different machine learning methods. First of all Linear Regression and Random Forest will be used to determine which ESG features best explain financial performance, and in this way determine which ESG factors have the most impact on financial outcomes. To determine which ESG factors have the most influence on the financial performance, feature selection will be applied in combination with both models. This will be done with the use Recursive Feature Selection.

In addition to these methods, this research will explore two approaches for predicting financial performance, namely Long Short-Term Memory (LSTM) and AutoRegressive Integrated Moving Average (ARIMA). By including exogenous variables into these models it is investigated whether incorporating ESG features enhances the accuracy of financial performance predictions. First the two models will be created without the inclusion of exogenous variables, after which they will be compared to the models including exogenous variables.

To be able to determine which exogenous variables will be included, four different scenarios will be created. In the first scenario the exogenous variable is based on the correlation between the target variable and the ESG features. The second scenario will base its exogenous variable on the outcomes of the Granger Causality test. The Third scenario will base the exogenous variable on the features found important by the Linear Regression and Random Forest model. The last scenario is a benchmark scenario, using 'ESG Combined Score' as exogenous variable. Both models, with and without exogenous variables, will be optimized by tuning the hyperparameters. The tuning of the hyperparameters will be done with the use of Grid Search. In this way the models with and without exogenous variables can be compared.

For all methods not only different ESG features will be investigated, but also different financial indicators. This will be done for Earnings per Share (EPS), Return on Equity

(ROE), and Return on Assets (ROA). Including these different financial indicators allows for a deeper understanding of the relationship between ESG performance and financial outcomes. Each of these indicators offers unique insights into different aspects of a company's financial well-being.

After applying these methods, a conclusion is drawn about which ESG feature mostly impact financial performance, and which machine learning methods can best be used to quantify this relationship.

1.4 Relevance for Host Company

This research hold profound relevance to KPMG as we aim to specialize in providing ESG solutions and advice to our clients. KPMG's top priority's worldwide are technology, ESG and talent. Because of this KPMG is invested in sustainability.

The ESG solutions created by KPMG are developed to be able to give clients more insights, and in this way meet laws and regulations concerning ESG. KPMG is dedicated to offering ESG solutions and advisory services that not only help our clients meet regulatory requirements but also enhance their overall sustainability practices. This not only strengthens the company's brand name and values but demonstrates its commitment to staying at the forefront of ESG discussions. This goes beyond compliance with laws and regulations, attracting clients and stakeholders who value a comprehensive approach to ESG considerations.

The findings from this research will directly inform the development and refinement of our ESG solutions and advisory services, ensuring they align not only with regulatory standards but also with the financial success of our clients.

1.5 Thesis Outline

This thesis is structured as follows. In chapter 2 'Background' the most important returning terms to understand this research will be explained. Next to that some previous research on the topic will be discussed. In chapter 3 'Data Analysis' the used database and obtained data is described. After this the process of the cleaning of the data will be explained, and feature engineering will be discussed. Lastly, the used feature selection method will be explained. Chapter 4 'Methods' describes the machine learning methods that are applied in the research. Other than that the theory around the determination of exogenous variables is explained, the way the parameters are tuned, and the methods used to validate the models. Chapter 5 'Results' describes the outcomes of the machine learning methods, and the obtained results. In chapter 6 'Conclusion', a conclusion is drawn based on the outcomes of the 'Results' chapter. Lastly, chapter 7 'Discussion and Future Research' discusses the limitations in the research, and proposes ways to improve the research in the future. All of the code written for this thesis can be found at `https://github.com/NinaDePiena/BA_master_thesis`.

2

Background

This chapter provides information on commonly used terms which are crucial to understand the remainder of the research. This includes the term ESG (Social, Environmental, Governance) and the different financial indicators.

2.1 ESG (Environmental, Social, Governance)

ESG is an acronym for Environmental, Social, Governance. It takes into account the environmental impact, social responsibility, and corporate governance practices of companies to evaluate their sustainability and ethical performance. The ESG principle was formally proposed in 2004, and ever since it has been actively practiced in Europe, America, and other developed countries (8).

Climate change, biodiversity, energy consumption, pollution, and waste management are examples of environmental factors that have an impact on the state and functionality of natural systems and the natural environment. Social variables include topics such as (in)equality, health, inclusivity, labor relations, workplace health and safety, human capital, and communities. Social factors are related to the rights, well-being, and interests of individuals and communities. In addition to executive leadership, executive compensation, audits, internal controls, tax evasion, board independence, shareholder rights, corruption, and bribery, governance factors also include how companies or entities incorporate social and environmental factors into their policies and processes (9).

ESG is flourishing and there is a growing interest in the implementation of ESG principles worldwide (2). Even though ESG research has transitioned from a niche concept to a widespread investment strategy, it still faces a lot of challenges. There is a lack of standardized ESG metrics and reporting frameworks, which leads to inconsistency in data quality and comparability (7). Other than that there is limited research on the long term effects of ESG integration, simply because of the limited data and information (2).

Different databases may use different ways to evaluate companies, which makes ESG scores somewhat subjective. For this research ESG scores and ESG related factors from the database LSEG (previously Refinitiv) are used. All of these scores and factors will be explained in chapter 3. The reason to use this database is that it is quite extensive. Is contains data on all required financial indicators, as well as a lot of ESG data on different categories. Other than that it is a database which is accessible with a KPMG account, which made it easy to use.

2.1.1 ESG laws and regulations

Worldwide there is a lack of reporting requirements (7), and also laws and regulations on ESG are not consisted around the world. *'The ESG reporting landscape is dynamic, fragmented, and evolving.'* (7) That is why it is important to take into account that companies located at different locations might have different driving force behind their ESG performance. It might also be that they have to comply to different laws and regulations, which makes that their behavior around ESG might be different than the behavior from companies located at other countries.

2.2 Financial Indicators

To be able to measure the relationship between ESG performance and financial performance three financial indicators were chosen. These indicators were chosen based on the extent to which they can express financial performance. For this research the financial indicators Earnings per Share (EPS), Return on Equity (ROE) and Return on Assets (ROA) are used. The construction and added value of each of these indicators will be explained in the next subsections.

2.2.1 Earnings per Share (EPS)

As shown in equation 2.1 the Earnings per Share (EPS) is calculated as a company's net profit divided by the number of outstanding shares of its common stock. It indicates how much money a company makes for each share of its stock, and serves as an indicators of a company's profitability (10). *'The higher the earnings per share, the better the performance and profitability of the firm is deemed to be'* (11), this is the reason to use this indicator as one of the financial indicators for this research.

$$\text{Earning Per Share (EPS)} = \frac{\text{Net Profit}}{\text{Number of Outstanding Shares of Common Stock}} \tag{2.1}$$

2.2.2 Return on Equity (ROE)

The Return on Equity (ROE) is calculated as the net income of a company divided by the average stockholders' equity, which can be found in equation 2.2. The return on equity determines the ability of a company to generate revenue from the capital it has invested. The invested capital, which is used to generate profit, originates from investors' investments in the company's stock and retained earnings. The better the company is at generating profit from its investment, the higher the return (10). *'Return on Equity is a commonly used analysis by investors and corporate leaders, to measure how much profit can be the right owner's own capital. For investors, the analysis of return on equity is important because the analysis can determine the benefits of the investments made. For companies, this analysis is important because it is a pull factor for investors to invest'* (11). This is the reason to use this financial indicator in this research.

$$\text{Return On Equity (ROE)} = \frac{\text{Net Income}}{\text{Average Total Equity}} \tag{2.2}$$

2.2.3 Return on Assets (ROA)

The Return on Assets (ROA) is calculated as the net income of a company divided by the average total assets, which can be found in the equation 2.3. *'The return on assets is one of the most widely used profitability ratios because it is related to profit margin and assets turnover, and shows the rate of return for creditors and investors of the company'* (12). The return on assets measures how well companies can use their assets to successfully generate profit. The higher the return, the more profit is created from asset use (10). It is stated that *'The return on assets are positively and significantly related to firm value'*

(13), which makes it a good financial indicator for this research.

$$\text{Return On Assets (ROA)} = \frac{\text{Net Income}}{\text{Average Total Assets}} \tag{2.3}$$

2.3 ESG and Financial Performance

There are already various researches done on the relationship between ESG indicators and financial performance. One of this is the research by Friedea , Buschb and Bassenb (2015) (5), titled 'ESG and financial performance: aggregated evidence from more than 2000 empirical studies'. In this research the relationship between ESG criteria and corporate financial performance (CFP) is investigated. This is done by exploring over 2000 studies between the 1970s until 2015. It describes how in approximately 90% of the investigated studies a nonnegative ESG–CFP relation was found. It was also stated that since the mid-1990s, the positive correlation patterns in primary studies have been stable over time. Based on this sample, they clearly found evidence for the business case for ESG investing.

Another study is the study by Whelan, Atz and Clark (2021) (14), called 'ESG AND FINANCIAL PERFORMANCE'. Within this research they investigated over a 1000 research papers between 2015 and 2020 on the relationship between ESG and financial performance. *'There was found a positive relationship between ESG and financial performance for 58% of the "corporate" studies focused on operational metrics such as ROE, ROA, and only 8% showing a negative relationship'* (14). This indicates that the use of ROE and ROA may be good indicators to measure financial performance based on ESG performance.

Zhao, Guo, Yuan, Wu, Li , Zhou and Kang (2018) (6) did a research on the relationship between ESG and financial performance in the power market in the paper 'ESG and Corporate Financial Performance: Empirical Evidence from China's Listed Power Generation Companies'. Here again a positive effect between ESG and financial performance was found. *'The results show that good ESG performance can indeed improve financial performance, which has significant meanings for investors, company management, decisionmakers, and industry regulators'* (6). It was concluded that by improving ESG the corresponding financial performance could also be improved.

Research by Almeyda, and Darmansyah (2019) (2), 'The Influence of Environmental, Social, and Governance (ESG) Disclosure on Firm Financial Performance' revealed that there is a substantial impact of ESG disclosure overall on the financial performance of real estate companies as determined by accounting indicators like ROA and Rate of Change (ROC). The research was based on a period of five years, and used data from 77 listed real estate companies. *'The study results indicated that there is a statistically significant positive relationship between Environmental disclosure and firm ROC as well as Stock Price. Other than that it also revealed that there's no significant influence of the Social and Governance pillar factor on firm financial performance'.* (2)

3

Data Analysis

This chapter describes the Data Analysis process. First a description of the data is given, which is obtained from the LSEG database. After this the cleaning of the data is explained, including handling missing values and outliers. This is followed by the exploration of feature engineering, to create new variables. The chapter concludes with feature selection, which is used to identify the most important ESG variables.

3.1 Data Description

The data used in this research comes from the LSEG database, formerly known as Refinitiv. This database contains data for the required financial indicators, as well as data on the ESG variables. Since this research focuses on the relationship between ESG performance and financial performance of companies that are clients of KPMG, data from different KPMG clients was considered. This resulted in the data from Microsoft being used for this research. The reason for this is that Microsoft is a close client of KPMG, and the LSEG database had sufficient ESG data on Microsoft.

3.1.1 ESG variables

The LSEG (London Stock Exchange Group) ESG database is a comprehensive platform that houses ESG scores for over 15,500 public and private companies globally. These scores are designed to assess and benchmark companies' performance on key sustainability metrics. Essentially, the ESG scores rely on publicly-reported data to measure a company's ESG performance.

The database contains 10 different ESG categories namely; resource use, emissions, innovation, workforce, human rights, community, product responsibility, management, share-

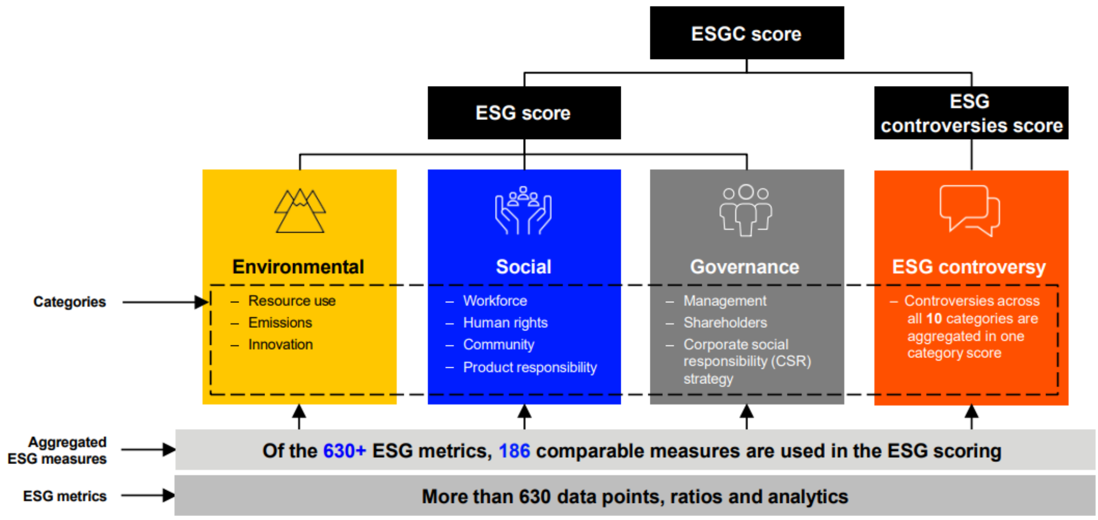

Figure 3.1: Construction of ESG Scores: A Visual Overview (figure obtained from the LSEG ESG documentation)

holders and corporate social responsibility. These individual category scores are combined into three overall pillar scores: environmental, social, and corporate governance. The ESG pillar score represents a relative sum of category weights, with variations in weights across industries.

Next to the category scores, pillar scores and ESG score, there is also a controversies score and a ESGC score. The controversies score provides a numerical indicator of how ESG controversies impact a company's overall sustainability performance. The score is build out of 23 ESG controversies topics. Companies with no controversies receive a score of 100, while those with recent controversies are penalized based on the severity of the issues. Lastly there is the ESGC scores, which is the overall ESG score combined with the controversies score. The ESGC score is calculated as the average of the ESG score and ESG controversies score when there are controversies during the fiscal year. The ESG score is equal to the ESGC scores in the case the controversies score is greater than the ESG score

The overview of all different scores and how they are connected can be found in figure 3.1. Next to that a table with all scores and their explanation can be found in table 7.1. All of these features are numerical scores ranging from 0 to 100.

The scoring methodology uses a percentile rank approach to guarantee robustness by reducing sensitivity to outliers. This method considers three possibilities: the number of companies performing worse, the number of companies having the same value, and the number of companies having a value at all. Scores are updated weekly, this means that historical records can also still change. This goes back to a maximum of 5 years. Scores older than 5 years will be marked 'definitive', and will not be changed anymore, even if there are updates to the underlying data. The higher the score the better the relative ESG performance and the higher the degree of transparency in ESG reporting.

To learn more about how the different ESG scores are determined, one may refer to the 'Environmental, social and governance scores from LSEG' documentation. This documentation is openly available and explains in detail how all ESG scores are determined by the LSEG database team.

Next to the previous mentioned scores, some other factors are included based on discussions with the KPMG sustainability team. From meetings with the sustainability team it was concluded that there are 5 topics on which there is specific focus from KPMG. These topics are Carbon, Human Rights, Circular Economy, Biodiversity, and Pollution. The LSEG database was investigated based on these topics to see if some additional variables should be included. This resulted in adding a few more variables, which can be found in table 7.2. In this table again the scores are numerical values ranging from 0 to 100. The 'CO2 Equivalent Emissions Total' is also a numerical feature expressing the amount of emission that year. The 'Human Rights Controversies' is a numerical feature expressing the number of controversies in that year. All the remaining features in the table are expressed with either true or false. Only for the pollution topic no additional factors are included due to lack of data on this topic within the database.

3.1.2 Financial Indicators

Next to the ESG related variables, also the financial indicator variables can be found in the LSEG database. The LSEG database includes the EPS, for which the definition from the database of the column can be found below.

Earnings per share – actual: *'The company's actual value normalized to reflect the I/B/E/S default currency and corporate actions (e.g. stock splits). Earnings Per Share is defined as the EPS that the contributing analyst considers to be that with which to value a security. This figure may include or exclude certain items depending on the contributing analyst's specific model.'* (15)

The LSEG database also contains columns for the ROE and the ROA. For both the definition of the database can be found below.

Return on Equity – Actual: *'The company's actual value normalized to reflect the I/B/E/S default currency and corporate actions (e.g. stock splits). Return On Equity is a profitability ratio calculated by dividing a company's net income by total equity of common shares.'* (15)

Return on Assets – Actual : *'The company's actual value normalized to reflect the I/B/E/S default currency and corporate actions (e.g. stock splits). Return On Assets is a profitability ratio and as such gauges the return on investment of a company. Specifically, ROA measures a company's operating efficiency regardless of its financial structure (in particular, without regard to the degree of leverage a company uses) and is calculated by dividing a company's net income prior to financing costs by total assets.'* (15)

3.2 Data Cleaning

After obtaining the data the first step was to investigate and clean the data. First of all the data was modified by reversing it. This was done because the initial datapoints go back in time, but the goal for this research is to predict future values. Other than that the data was formatted into time series data to be able to work with it. To be able to align the quarterly financial data with the yearly ESG data, a very strong assumption was made that the ESG performance followed a linear trend. I this way the financial and ESG data was matched to be the same frequency.

The visual investigating was done by creating plots, and by looking into the summary statistics. In figs. 7.1 to 7.4 some scatterplots for the financial data, and for a few of the ESG features can be found. When looking at the numerical summary it can be found that most values range from 0 to 100, which are the scores. Other than that there are a few features that only take 0 and 1 values, indicating if something happened or not, for example 'Eco-Design Products'. Lastly the 'CO2 Equivalent Emission Total' takes values between 1.526000e+05 and 6.520660e+06 and has a mean of 2.155072e+06.

3.2.1 Missing Values

In tables 7.3 and 7.4 the missing values per column can be found. The missing values were handled by applying the interpolate function in python from the pandas library, with the linear method. This function handles missing values by creating a line between the nearest data points without missing values, and uses this line to estimate the missing values between these points. After applying this method a few columns still contained some missing values. These missing values where found either at the beginning or at the end of each column and could not be handled by the interpolation method. Since the number of remaining missing values was quite small the decision was made to not remove these rows, but to fill this missing values with the use of two different methods. This to make sure no important information from the other complete columns would be deleted. First the forward fill function was applied, and after that the backward fill function. In this way the missing values at the beginning and end of each column would be filled with the values located closest to it. The reason to base the missing values on neighboring values has to do with the fact that the data is time series. This makes that neighboring values are more explanatory for the missing values than the other values in the dataset. Because of this taking for example the mean of the whole dataset to fill in missing values would give a distorted image and would not be an appropriate way to fill in missing values.

3.2.2 Outliers

Lastly it was checked if the data contained any outliers. First of all the data was investigated on impossible values. It was checked if all scores were within the score range of 0 to 100. This was the case so based on this information no outliers were detected within the scores features. The other feature values also lay within a realistic domain, with no impossible values. Other than just checking manually and visually for realistic values, a outlier detection method was applied to see if any outliers where present which were less obvious. This was done by applying the Local Outlier Factor method. The Local Outlier Factor method works by evaluating the local density properties of each data point in relation to its neighbors, and determines based on this information if a datapoint is an outlier. *'The outlier factor is local in the sense that only a restricted neighborhood of each object is taken into account'* (16). The reason to choose this method and not a simple-distance based method is because there is a lot of variability in the data, and the neighboring datapoint are more explanatory than the other datapoints. The method returned a few datapoint which it considered to be outliers. Although the Local Outlier Factor method appointed some points to be outliers, the decision was made to not remove these points. This had to do with the fact that the 'outliers' did not seem like outliers and seemed quite realistic within the course of the data. To prevent loosing crucial information they remained within the dataset. An example of such a determined outlier can be found in graph 3.2. Based on both visual investigation and applying the Local Outlier Factor method, this resulted in not removing any outliers from the data.

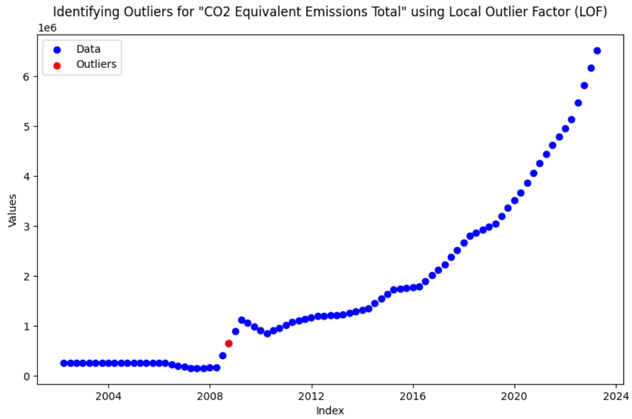

Figure 3.2: Potential outlier from Local Outlier Factor Method

3.3 Feature Engineering

Feature engineering is the process of extracting features from unprocessed data and converting them into formats that can be used by machine learning models. It serves as a crucial step in the process of machine learning as it can make the modeling easier, and in this way the process can generate outputs of higher quality (17). *'Feature engineering is the process of formulating the most appropriate features given the data, the model, and the task'* (17). For this research a few feature engineering methods are applied which will be explained in the next paragraphs.

3.3.1 Time Domain Features

The first features created are time domain features, or temporal features in the time domain. Temporal features in the time domain represent summaries of relevant historically observed values (18). This could for example be taking the minimum, maximum or mean over a certain time window. In this research the newly created temporal features were based on the mean over a certain time window. This was done for three different time windows, namely 2, 3 and 10. Doing this helps smoothing the feature and capturing trends. It provides a more general representation of the feature, which could make it easier for the models the learn patterns. The larger the time window the more smooth the new feature.

By reducing the complexity of the feature overfitting can be prevented and the underlying essence of the feature can be captured.

3.3.2 Delta Features

The next few newly created features are delta features. Delta features are features focused on the change over time between two time points (19). To be able to capture this change, features are created that express the percentage change, the absolute difference and the change expressed in a categorical way. The percentage change expresses the relative change in the timeseries between two consecutive values. The difference expresses the absolute difference in value between two consecutive values. Lastly the categorial change expresses the change between two consecutive values in a categorical way. This is done by labeling it to be ‘increasing’, ‘decreasing’, or ‘same’. To be able to use this categorial feature in the models it is made into a numerical value with the possible values being 1, -1 or 0. The percentage change and absolute difference have different advantages and disadvantages. Percentage change can be sensitive for small initial values, but is very useful for normalization and relative comparison. Absolute difference on the other hand lacks normalization, but it does offer an more straightforward interpretation. For this reason both methods are considered.

3.3.3 Lagged Features

To be able to capture historical dependencies in the data, lagged features are created. A lagged feature is based on its original feature but shifted back in time by a certain ‘lag’. In this way past observations are incorporated as features, which could possibly reveal temporal relationships and patterns. Due to missing domain knowledge on ESG and lagged features, the time lags were chosen randomly. Lagged features were created with a lag of 1, 2 and 3. This means the data is shifted for 1, 2 and 3 quarters of the year.

3.3.4 Binning Features

Binning features were added to be able to simplify features by dividing the values of a feature into discrete bins. *‘With fixed-width binning, each bin contains a specific numeric range. The ranges can be custom designed or automatically segmented, and they can be linearly scaled or exponentially scaled’* (17). The binning of features was done for all of the scores features, and for the ’CO2 Equivalent Emissions Total’ feature. In the case of all features that are scores, the number of bins is 10, with also a size of 10. This results in

10 bins with all the size of 10, and could be seen as a grading system from 1 to 10. The 'CO2 Equivalent Emissions Total' feature was divided into 100.000 bins, with all a size of 100, resulting in a maximum value of 10.000.000.

3.3.5 Interaction Features

'An easy way to extend the linear model is to include combinations of pairs of input features. This allows us to capture interactions between features, and hence these pairs are called interaction features' (17). This was done by creating interaction features for every pair of pillar scores. The ESG score already includes the interaction between all three pillar scores, while these newly created interaction features only look at the interaction between two out of the three features. Next to that, some interaction features based on the binary variables were created. These feature were created by combining the binary variables with the pillar scores. This was for example done for the 'Human Rights Policy' - which indicates if at that point of time there was a policy for human rights -, and the social pillar score. The last interaction feature created was a feature based on the interaction between the environmental pillar score and the 'CO2 Equivalent Emissions Total'. This new feature can possibly provide more information on the environmental impact.

3.3.6 Scaling Features

Lastly all of the features were scaled. In the case that a model is sensitive to the scale of the input features, scaling the features may help (17). Since both Linear Regression and Neural Networks (which LSTM is a part of) are sensitive to the scale of the input features (20)(21), the decision was made to scale the features to avoid a bias. Since most of the features within the dataset are scores within a range of 0 to 100, that range was also chosen to be the range for the scaled features. This makes the features contribute equally to the analysis.

The newly engineered features can be found in table 7.5. Where before the dataset contained 26 features, after feature engineering this ended up being 280 features. To be able to bring this back to a useful dataset for the machine learning part, feature selection was applied which will be discussed in the next chapter.

3.4 Feature Selection

The goal of feature selection is to identify the most relevant features within a dataset. By eliminating features that are not relevant, feature selection improves model performance

(22). The two main advantages of applying features selection are better interpretability, and higher accuracy which then avoids the risk of overfitting (23).

3.4.1 Recursive Feature Elimination (RFE)

The RFE selection method is a feature selection method which ranks features based on their importance with the use of a recursive process. The feature importance is measures at each iteration, and the least relevant feature is removed (23). *'The recursion is needed because for some measures the relative importance of each feature can change substantially when evaluated over a different subset of features during the stepwise elimination process (in particular for highly correlated features)'* (24). This feature selection method is applied in combination with a model, and the importance of the feature is based on how well it works within the model. *'The feature selection process itself consists only in taking the first n features from this ranking'* (24). In this research the RFE was applied in combination with Linear Regression model and the Random Forest model, which will be explained in chapter 4. The features found to be most important for each of the methods after applying RFE will be explained in chapter 5

4

Methods

This chapter discusses the methods used to investigate the relationship between ESG performance and financial performance. The methods chosen for this study are Linear Regression, Random Forest, LSTM (Long Short-Term Memory), and ARIMA (AutoRegressive Integrated Moving Average). Each of these methods will be explained in detail in the following sections.

4.1 Linear Regression

Regression models are models widely used. Linear regression is a method to define the relationship between one dependent variable, and one or multiple independent variables (25). *'They are simple and often provide an adequate and interpretable description of how the inputs affect the output. For prediction purposes they can sometimes outperform fancier nonlinear models, especially in situations with small numbers of training cases, low signal-to-noise ratio or sparse data'* (26). In a Linear Regression model the variable that is predicted is called the response variable, while the variables that are used to predict the response variable are called the predictor variables (25). *'The linear model either assumes that the regression function E(Y|X) is linear, or that the linear model is a reasonable approximation'* (26). Since in this research one dependent variable (financial performance) is predicted based on multiple independent variables (different ESG scores), the Multiple Linear Regression model is used. The formula for the Multiple Linear Regression model can be found in equation 4.1.

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j \cdot X_j + \varepsilon \tag{4.1}$$

$$\begin{aligned} Y &: \text{Dependent variable} \\ X_1, X_2, \ldots, X_p &: \text{Independent variables} \\ \beta_0 &: \text{Intercept} \\ \beta_1, \beta_2, \ldots, \beta_p &: \text{Coefficients of the independent variables} \\ \varepsilon &: \text{Error term} \end{aligned}$$

We assume that our data consists of n observations: $(Y_1, X_{1,1}, \ldots, X_{1,p}), \ldots, (Y_n, X_{n,1}, \ldots, X_{n,p})$, where:

$$\begin{aligned} (Y_i, X_i) &: \text{The } i\text{-th observation in the sample} \\ Y_i &: \text{Dependent variable for the } i\text{-th observation} \\ X_i &: \text{Vector of predictors for the } i\text{-th observation} \\ X_i = (X_{i,1}, \ldots, X_{i,p}) &: \text{Predictor variables for the } i\text{-th observation} \end{aligned}$$

Inference on the parameters $\beta_0, \beta_1, \ldots, \beta_p$ is based on the sample data $(X_1, Y_1), \ldots, (X_n, Y_n)$, where each $X_i = (X_{i,1}, \ldots, X_{i,p})$.

The first coefficient in the formula (β_0) is called the intercept, while the other coefficients (β_1 through β_p) are called partial regression coefficients. *'In a model with more than one predictor, the slope represents the effect of a predictor when all other predictors are taken into account'* (27). A statistical model consists of two parts. There is the systematic part that expresses the part of the outcome that is related to X ($\beta_0 + \sum_{i=1}^{n} \beta_i \cdot X_i$), and next to that there is a random part which represent the variation, also the error (ε) (27). In the case of multiple regression the assumption is made that the errors are independent, that they follow a normal distribution, and that the variance of the residuals is constant (homoscedastic) (27). Here the residuals are the difference between the observed and the predicted values (27)(25), the formula for this can be found in equation 4.2. Note that these are different from the error.

$$e_i = Y_i - \hat{Y}_i \tag{4.2}$$

e_i : Residual for observation i

Y_i : Observed value of the dependent variable for observation i

$\hat{Y}_i$: Predicted value of Y for observation i

'A residual represents how wrong the regression prediction is for a given individual. The regression line can be constructed by minimizing the sum of the squared residuals' (27). To do this the Ordinary Least Squares method can be used. The Ordinary Least Squares method returns the coefficients that result in the smallest Residuals Sum of Squares (27). The formula for the Residual Sum of Squares can ve found in equation 4.3.

$$RSS = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{4.3}$$

RSS : Residual Sum of Squares

Y_i : Observed value of the dependent variable for observation i

$\hat{Y}_i$: Predicted value of Y for observation i

n : Number of observations

With the use of this formula the best linear fit to the data is found (26).

4.2 Random Forest

Introduced by L. Breiman in 2001, the random forest algorithm has been an extremely successful method for both classification and regression tasks (28). A random forest is a collection of multiple decision trees (28). The process of creating those different decision trees involves creating sample fractions of the dataset with the use of bootstrapping. which is a resampling techinque. After that a randomized tree predictor is trained on each subset, followed by the aggregation of these predictors (28). The combination of this random sampling from the dataset and aggregation technique is known as bagging (bootstrap-aggregating). It entails generating multiple bootstrap samples from the original dataset, constructing a predictor for each sample, and then averaging the predictions (28). A simplified image of a Random Forest model can be found in figure 4.1.

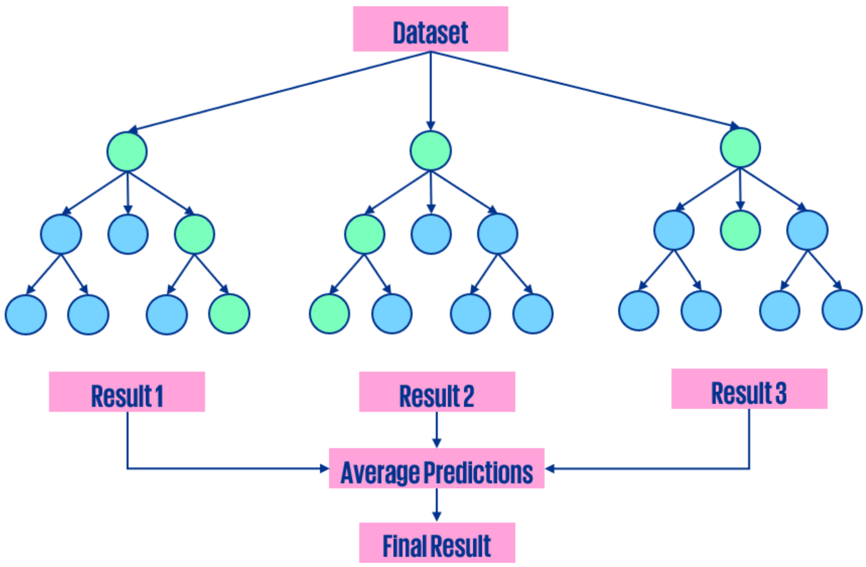

Figure 4.1: Simplified image of a Random Forest model

It is worth noting that the behavior of random forest can be reassured by the Strong Law of Large Numbers. This theory guarantees that random forests always converge, and will not overfit (29).

4.3 ARIMA(X)

ARIMA (AutoRegressive Integrated Moving Average) is a model used to analyze and forecast timeseries data. It consist of different building blocks among which the Autoregressive (AR) model and Moving Average (MA) model. Those two models together form the ARMA model which can be used to forecast stationary data. In the case the data is not stationary the ARIMA model can be applied. This is the ARMA model in combination with a differentiating term. Lastly there is the ARIMAX model, which is an ARIMA model including one or multiple exogenous variables. All of these different terms and the combination of terms will be explained in the next sections.

4.3.1 AR

AR stands for Auto Regressive. The Autoregressive model predicts the future value based on a linear combination of the historical values (30). The formula for an Autoregressive model can be found in equation 4.4.

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \epsilon_t \tag{4.4}$$

$$\begin{aligned} X_t &: \text{Value at time t} \\ c &: \text{A constant} \\ \phi_1, \phi_2, ..., \phi_p &: \text{the autoregressive parameters} \\ \epsilon_t &: \text{The white noise} \end{aligned}$$

4.3.2 MA

MA stands for Moving Average. The moving Average model predicts the future value based on previous forecasting errors (30). The formula for an Moving Average model can be found in equation 4.5.

$$X_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q} \tag{4.5}$$

$$\begin{aligned} X_t &: \text{Value at time t} \\ \mu &: \text{The mean of the time series} \\ \theta_1, \theta_2, ..., \theta_q &: \text{the moving average parameters} \\ \epsilon_t &: \text{The white noise} \end{aligned}$$

4.3.3 ARMA

The ARMA model is a combination of the AR and the MA model. It predicts future values based on both past values, and past forecasting errors. *'An ARMA model combines the ideas of AR and MA models into a compact form so that the number of parameters used is kept small, achieving parsimony in parameterization'* (31). The formula for the general ARMA(p,q) model can be found in 4.6. Where ϵ_t is a white noise series and p and q are nonnegative integers (31).

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q} \tag{4.6}$$

$$X_t : \text{Value at time t}$$
$$c : \text{A constant}$$
$$\phi_1, \phi_2, ..., \phi_p : \text{the moving average parameters}$$
$$\theta_1, \theta_2, ..., \theta_q : \text{the moving average parameters}$$
$$\epsilon_t : \text{The white noise}$$

4.3.4 ARIMA

The ARMA model becomes an ARIMA model in the case a non-stationary factor is added to the model. ARIMA is a combination of the AR and the MA models which have undergone a differencing process, for which the I stands in ARIMA (30). *'An ARIMA Model corresponds to ARMA after finitely many times differences the data'* (32). The order d in ARIMA(p,d,q) indicates how many differences are needed for the data to become stationary (32). The value for d can be found by differencing the data until it is stationary, after which the values for p and q should be found. The degree of p can be found from the plot of the partial autocorrelation function (PACF) and the degree of q can be found from the autocorrelation function (ACF) (30).

4.3.5 ARIMAX

Next to the ARIMA model there is the ARIMAX model, which stands for Autoregressive Integrated Moving Average with Exogenous Variables. This model is an extension of ARIMA where independent variables are added, also known as exogenous variables (33). The ARIMAX model can measure the influence of external factors (the exogenous variables) on the dependent variable (the endogenous variable) (34). *'ARIMAX is helpful since it combines the time series and regression components into one model'* (33). The ARIMAX model is a model combining a multiple linear regression model with autoregressive (AR) terms and moving average (MA) terms (34). The formula of ARIMAX can be found in 4.7.

$$X_t = \beta_0+\beta_1 Y_{1,t}+\beta_2 Y_{2,t}+...+\beta_q Y_{q,t}+\phi_1 X_{t-1}+\phi_2 X_{t-2}+...+\phi_p X_{t-p}+\theta_1\epsilon_{t-1}+\theta_2\epsilon_{t-2}+...+\theta_q\epsilon_{t-q}+\epsilon_t \tag{4.7}$$

$$X_t : \text{Value at time t}$$
$$Y_{1,t}, Y_{2,t}...Y_{q,t} : \text{the exogenous variables at time t}$$
$$\beta_0, \beta_1...\beta_q : \text{the coefficients for the exogenous variables}$$
$$\phi_1, \phi_2, ..., \phi_p : \text{the moving average parameters}$$
$$\theta_1, \theta_2, ..., \theta_q : \text{the moving average parameters}$$
$$\epsilon_t : \text{The white noise}$$

4.4 LSTM

LSTM (Long Short-Term Memory) is a variant of the RNN (Recurrent Neural Network), and just like ARIMA used to analyze and forecast timeseries data. LSTM specialize in capturing long-term dependencies in sequential data. Unlike traditional RNNs, LSTM uses memory cells and gates to be able to learn more effectively. LSTMs are know to be able to capture complex patterns well. In the next sections the traditional RNN as well as the LSTM will be explained.

4.4.1 RNN

Recurrent Neural Networks (RNN) are very suitable for processing time series data (35). A RNN is a Neural Network consisting of input, hidden and output nodes (36). A simplified version of a RNN can be found in 4.2. The output of the RNN depends on previous computations and in this way the RNN is able to develop memory, which is encoded in the hidden state nodes (36). *'RNNs are learning machines that recursively compute new states by applying transfer functions to previous states and inputs'*(36). However, RNN are sensitive to only short-term information, which makes it difficult to model long-term dependencies (37). This limitation can be solved by using Long Short-Term Memory (LSTM), which the remainder of this section will be about.

4.4.2 LSTM

Recurrent Neural Networks can combine several temporal features, which makes them well suited for handling time series data (35). However, RNNs face challenges such as gradient disappearance and gradient explosion, particularly when the number of time steps in the network is either too small or too large (35). In 1997 Hochreater and Schmidhuber proposed

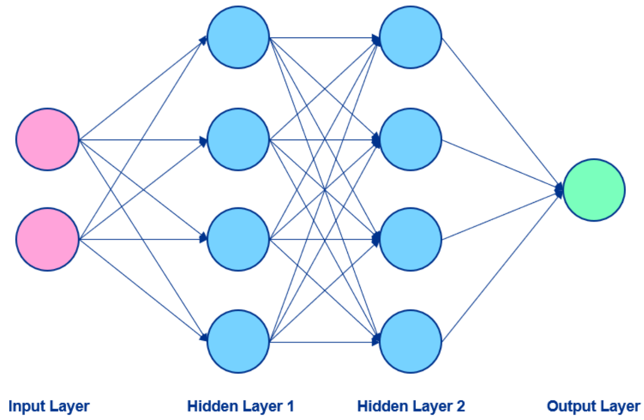

Figure 4.2: Simplified image of a Recurrent Neural Network

a method which is an improvement on the RNN model, named Long Short Term Memory (LSTM) (38). *'LSTM is widely used nowadays due to its superior performance in accurately modeling both short- and long-term dependencies in data'* (36). The difference between a RNN and LSTM is that LSTM works with a so called cell, which acts as a memory unit (36). Within this memory unit three gates can be found; the Input Gate, Forget Gate, and Output Gate (39). A simplified overview of the LSTM model can be found in 4.3. *'The Long Short-Term Memory (LSTM) proposed by Hochreiter and Schmidhuber can solve its gradient disappearance and gradient explosion problems well by adding input gate, forget gate, and output gate to RNN'* (35). The gates in the LSTM model are used to be able to control the flow of information to reach the cell. By selectively passing the information and in this way avoiding information overload, the problem of gradient disappearance and gradient explosion is handled (39). To be able to control the amount of information being passed on through the gates, a Sigmoid function is used. The output of this Sigmoid function is a number between 0 and 1, indicating the degree of the information kept (39). *'LSTM tries to solve the vanishing gradient problem by not imposing any bias toward recent observations, but it keeps constant error flowing back through time'* (36).

4.4.3 LSTM with exogenous variable

To potentially improve the forecast made by the LSTM model it is possible to incorporate one or multiple exogenous variables. The idea is that these exogenous variables would have an effect on the endogenous variable, and in this way could potentially help the model make better predictions. In the case of this research this is done by incorporating ESG features

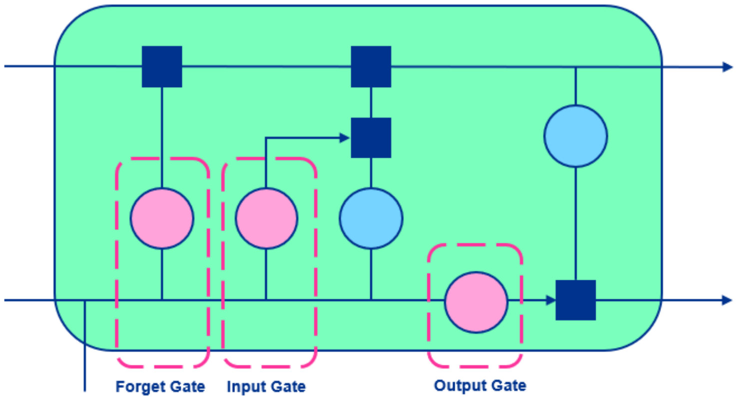

Figure 4.3: LSTM structure

into the LSTM model, to see if they have an effect on the accuracy of the prediction of the financial target variable.

4.5 Practical considerations of ARIMA and LSTM

ARIMA and LSTM are prominent techniques in time series forecasting, each with their own advantages and disadvantages. The first difference between the models is the required dataset size to be able to perform well. ARIMA is said to need at least 50 observations to be able to perform well, while for LSTM this ideally should be at least 200 (40)(41). An advantage of ARIMA is that it has a relativity simple model structure, and that it is easy interpretable (42). In terms of short term forecasting they have proven to often outperform more complex models (43). ARIMA also comes with some disadvantages, it relies on historical data, and because of this not very effective at predicting turning points in the data (43). Other than that it is a model which might commonly overfit during the identification stage (43). An advantage of LSTM is that it is quite suitable for capturing more complex patterns in the data (39). Next to that it has the ability to learn long-term dependencies, and it can handle the problem of gradient vanishing an exploding (37). A drawback that comes with the model is that it needs a lot of processing power (37).

In summary, the choice between ARIMA and LSTM depends on factors such as interpretability, model complexity, and the nature of the data. ARIMA is simple and easy to interpretate, while LSTM is more suitable to learn long term dependencies and complex patterns.

4.6 Granger Causality Test

To be able to determine good exogenous variables for both the ARIMAX and the LSTM models, the Granger Causality test is used. This is a statistical test to determine if one time series (the exogenous variable) can help predict another time series (the endogenous variable). X is said to Granger cause Y if historical information on X helps predict Y better than just information on Y by itself (44). Is it important that only X (in this case the ESG performance) Granger-causes Y (the financial performance), and not the other way around. This is because for X to be an exogenous variable to Y, Y must fail to Granger cause X (45). The Granger Causality test can use an F-test to see whether X Granger causes Y (44), which is also the method used in this research. The Null hypothesis of the test is that X does not Granger cause Y (46). When rejecting the Null hypothesis, this might indicate that X can help predict Y, which suggest Granger Causality.

4.7 (Hyper)parameter Tuning

To be able to tune the hyperparameters of the models the Grid Search method was used. Grid search considered all possible combinations of parameters within a given discrete parameter space (47). By trying all different combinations of parameters, an optimal set of parameters is found. A disadvantage of Grid Search is that it might miss the actual optimal solution, since only the specific grid values are considered, and not the combinations of values that lay between these points.

4.8 Evaluation Metric

The evaluation metric used for this research is the Root Mean Squared Error (RMSE). This metric measures the difference between the actual values of the dependent variable and the values predicted by the model (40). The formula for the RMSE can be found in equation 4.8. The RMSE is very sensitive to outliers due to the quadratic part of the formula (42). By squaring the error, high errors are penalized more, which means they have a bigger impact on the metric.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{4.8}$$

n : Number of data points

y_i : Actual observed value at data point i

$\hat{y}_i$: Predicted value at data point i

4.9 Validation of the models

To be able to validate the outcomes of the models and determine whether the performance of the models with exogenous variable is significant compared to the performance of the models without exogenous variables, k-fold cross validation is applied in combination with a permutation test.

K-fold cross validation is a widely used methods to evaluate a model and estimate the prediction error (48). *'The KCV consists in splitting a dataset into k subsets; then, iteratively, some of them are used to learn the model, while the others are exploited to assess its performance'* (49). Reasons to apply k-fold cross validation are more reliable results by reducing the variance of the estimation, and minimizing overfitting (47)(50). Other than that k-fold cross validation is useful in the case of a limited data set, which is the case in this research. K-fold cross validation allows each of the datapoints to be used both during training and validation, and in this way the models is able to increase its learning ability.

Since in this research the data is time series data, traditional k-fold cross validation cannot be applied due to the temporal nature of the time series data. For the model evaluation of time series data it is important to take the order of the datapoints into account, which is not done in the case of the traditional k-fold cross validation, where the data is randomly shuffled. 'TimeSeriesSplit' from scikit-learn is a k-fold cross validation method designed for time series data. This method, unlike the traditional method, takes the temporal order of the data into account. TimeSeriesSplit works in the way that the data is split into k consecutive folds, where each fold uses the past data for training, and the future data for the validation (51). A visualization of this process can be found in figure 4.4.

To be able to determine if there is a significant difference between the original models without exogenous variables and the models with exogenous variables, a statistical test is performed. A permutation test is applied to the arrays containing the RMSE's of the different folds. The reason to apply a permutation test, and not another statistical test, is because the permutation test does not make assumptions on the normality of the data, and works well on small datasets (52). A permutation test works by shuffling data to create random samples. After that it checks how often the result from these random sample is found to be as extreme as the actual results. If the observed result is very extreme compared to the random samples, the observed results is found to be statistically significant.

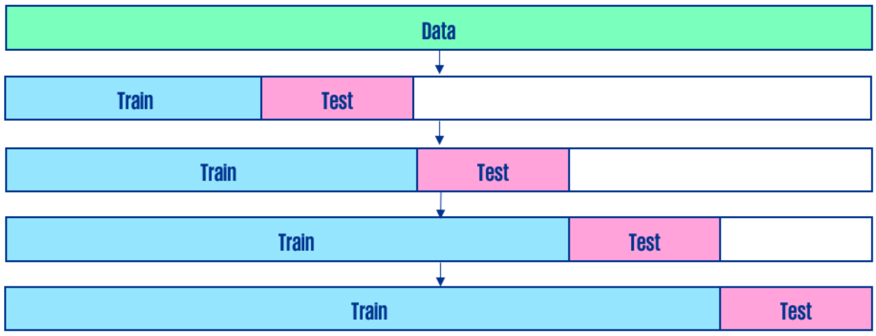

Figure 4.4: K-fold Cross Validation for Time Series

5

Results

In this chapter the results from this research will be discussed. In the first section the outcomes for the Linear Regression and the Random Forest models will be explained. Here, the most important ESG features will be determined. The next section will discuss the ARIMA(X) and LSTM models. Here the predictive power of both models will be compared with and without exogenous variables.

5.1 Linear Regression and Random Forest

Both Linear Regression and Random Forest are used to identify the most important ESG features. With the use of Recursive Feature Elimination (RFE) the optimal number of features for each model is determined, together with the feature importance of those features. This is done by by determining the RMSE for each number of features in the model, to identify the configuration that minimizes the RMSE. Due to computational costs the maximum number of features was set to be 30, which was around 10 percent of the total number of features. The RMSE's plotted against the corresponding number of features can be found in figs. 7.5 to 7.7 for Linear Regression, and in figs. 7.8 to 7.10 for Random Forest. It can be seen that in the case of Linear Regression the more features the lower the error. For Random Forest the RMSE will still change a lot while adding more features. The average RMSE for each model can be found in table 5.1. Here it can be seen that in all cases Random Forest outperforms Linear Regression. Next to that it can be found that the prediction for ROE and ROA outperforms the prediction for EPS for both methods. In figs. 7.11 to 7.16 the list of best features together with their importance can be found. When comparing the lists of most important features for both models it can be seen that the union of the list consist of the features found in table 7.6. And overview of the most important features for each model and the union of the models can be found in table 5.2.

	Linear Regression	Random Forest
EPS	0.148490	0.065044
ROE	0.052085	0.013144
ROA	0.018484	0.014035

Table 5.1: RMSE for Linear Regression and Random Forest

Financial Indicator	Linear Regression	Random Forest	Union
EPS	Governance Pillar Score_Mean_10	Human Rights Score_Mean_10	Human Rights Score_Mean_10
ROE	Governance Pillar Score_diff	ESG Combined Score_Mean_3	Human Rights Policy Score
ROA	Social Pillar Score	ESG Score_Mean_10	Eco-Design Products_Mean_10

Table 5.2: Overview of most important features

5.2 ARIMA(X) and LSTM

For both the ARIMA model and the LSTM model, four different scenarios are created. For each scenario the model without exogenous variable is compared to the model including an exogenous variable. The first scenario includes the best exogenous variable according to the correlation matrix. In table 7.8 the top 5 correlated variables for each of the financial indicators can be found. Scenario two includes the best exogenous variables based on the Granger Causality test. The top 5 exogenous variables based on the Granger Causality test can be found in table 7.7. An overview of the found exogenous variable based on either correlation or the Granger Causality test can be found in table 5.3. The third scenario includes the best exogenous variable according to the union of the outcomes of the Linear Regression and the Random Forest models. These variables can be found in table 7.6 . The last scenario is a benchmark scenario. Here the exogenous variable is the 'ESG Combined Score', which is the most general ESG score. General ESG scores were already found to have significant effect on the predictive power of the models in previous research. With the use of this benchmark scenario the quality of the choice of exogenous variables in the other scenarios can be determined, just like the quality of the models in general.

Financial Indicator	**Exogenous variable based on correlation**	**Exogenous variable based on Granger Causality test**	**Exogenous variable based on feature importance LR/RF**
EPS	CO2 Equivalent Emissions Total	CO2 Equivalent Emissions Total_Mean_10	Human Rights Score_Mean_10
ROE	Emission Score_Mean_3	Resource Use Score_bin	Human Rights Policy Score
ROA	Human Rights Breaches Contractor	Human Rights Breaches Contractor_Mean_3	Eco-Design Products_Mean_10

Table 5.3: Exogenous variables for each financial indicator based on correlation and Granger Causality test

5.2.1 ARIMA(X)

For an ARIMA model to work, the data should be able to become stationary after a certain number of differentiating. To check this an Augmented Dickey–Fuller test was applied. Here it was found that indeed after a certain number of differentiating the data could become stationary. The next step was to determine the initial values for p, d, and q. The values of d is the number of times the data need to be differentiated, and the values for p and q can be read from PACF and ACF plots. An example of the PACF and the ACF for one of the financial indicators (ROA) can be found in figure 5.1 and 5.2. The number where the cutoff is will determine the initial parameters. In the case of the ACF graph, the cutoff is at lag 7, which makes it AR(6). In the case of the PACF graph the cutoff is at 2 which makes it MA(1). The range of parameters values that the ARIMA(X) model will search in is based on these initial parameters. The range was determined by taking the initial p and q values, and by adding a few to that number to give the model the option to explore outside of the initial values.

After initializing the base model, the ARIMAX model including exogenous variables was created. To be able to validate the outcomes k-fold cross validation was applied with a split of 5. For both models in each fold hyperparameter tuning was applied by trying each combination of p, d and q in a range based on the initial parameters, as mentioned before. For each of the folds the RMSE of both models was saved. The average of the RMSE for all scenarios was calculated which can be found in 5.4. For each scenario the graph of the last fold together with the actual test data and predictions for both models can be found in figs. 7.17 to 7.20, 7.25 to 7.28 and 7.33 to 7.36.

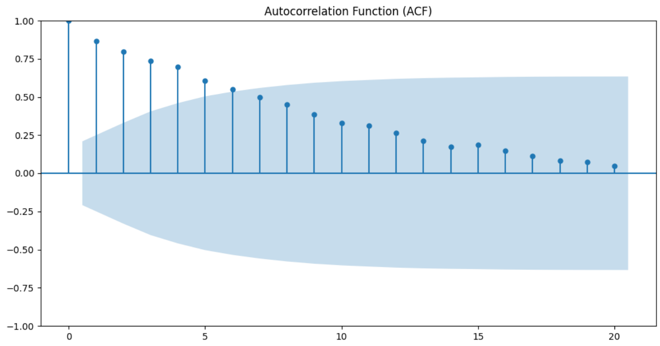

Figure 5.1: The ACF graph for ROA

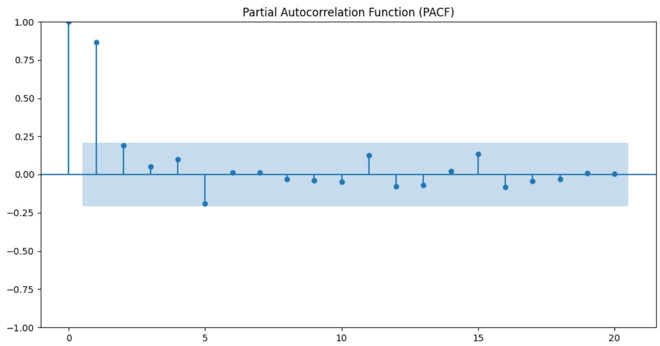

Figure 5.2: The PACF graph for ROA

5.2.2 LSTM

For the LSTM model a similar approach as for the ARIMA(X) model was applied. Again an initial model with and without exogenous variables was created, after which the data was split up in 5 folds. For both models hyperparameter tuning was applied with with use of Grid Search. For each of the folds the RMSE of both models was saved. The average of the RMSE for all scenarios was calculated which can be found in table 5.5. A graph of the last fold together with the actual test data and predictions for both models can be found in figs. 7.21 to 7.24, 7.29 to 7.32 and 7.37 to 7.40.

5.3 Comparison of ARIMA(X) and LSTM

An overview of the RMSE of all models can be found in tables 5.4 and 5.5. The RMSE of the model including an exogenous variable is highlighted in the case it outperformed the original model without exogenous variable. For the cases that the model including exogenous variable outperformed the original model, a significance test was performed to see if the difference between the original model and model with exogenous variable is indeed significant. For this a Permutation Test was used. The reason to use this test is that the Permutation test does not make any assumptions about the distribution of the data, which other tests normally do (52). It was found that in the cases the RMSE went down, this was not a significant result compared to the RMSE of the original model.

	ARIMA (corr)	**ARIMA (Granger)**	**ARIMA base**	**ARIMA combined**
Base Model EPS	0.143375	0.143375	0.143375	0.143375
Exog Model EPS	0.498119	1.261370	0.177613	0.141815
Base Model ROE	0.046278	0.046278	0.046278	0.046278
Exog Model ROE	0.051158	0.045293	0.642258	0.055767
Base Model ROA	0.028813	0.028813	0.028813	0.028813
Exog Model ROA	0.032556	0.033075	0.038839	0.028250

Table 5.4: Average RMSE for ARIMA models

	LSTM (corr)	LSTM (Granger)	LSTM base	LSTM combined
Base Model EPS	0.506180	0.454348	0.413332	0.632356
Exog Model EPS	0.356740	0.309037	0.485842	0.433764
Base Model ROE	0.068592	0.072119	0.121103	0.110796
Exog Model ROE	0.076717	0.126811	0.105054	0.083015
Base Model ROA	0.053618	0.056222	0.046488	0.052371
Exog Model ROA	0.046245	0.062546	0.058991	0.052102

Table 5.5: Average RMSE for LSTM models

6

Conclusion

This research was used to explore the relationship between ESG performance and financial performance with the use of machine learning. The main research question of this study was: 'What is the relationship between ESG performance and financial performance of a company, and how can machine learning models quantify this relationship?'

To be able to address this research question, several subquestions were formulated:

- Can ESG performance help predict financial performance more accurately?
- What ESG factors contribute to the prediction of financial performance?
- Which financial performance can be predicted based on ESG performance?
- Which selected method(s) works best to quantify the relationship between ESG and financial performance?

The following sections discuss the key findings of this research. Each of the sections discusses a subquestion and the findings, and the chapter ends with an overall conclusion and an answer to the research question. The term 'base models' in this chapter is referring to Linear Regression and Random Forest, while the term 'advanced models' is referring to LSTM and ARIMA.

6.1 Subquestion 1

Can ESG performance help predict financial performance more accurately?

It was found that in some cases adding an exogenous variable to the models reduced the RMSE. But after closer investigations and performing a significance test, it was found that the reduction in the errors was not significant for any of the models or any of the scenarios.

6.2 Subquestion 2

What ESG factors contribute to the prediction of financial performance?

6.2.1 Base Models

For both the Linear Regression and Random Forest the most important features were determined which can be found in figs. 7.11 to 7.16. Other than that the union of the most important features was determined which can be found in 7.6. An overview of the most important feature for each model and the union of the models can be found in table 5.2. What stands out is that a lot of the most important features are features either based on the overall ESG score, or on one of the three pillars. This is in line with previous research. Other than that variations of Human Rights features are found to be important for both ROE and EPS, and lastly Eco-Design Products_Mean_10 returned as the most important feature for the union of the models for ROA. Compared to previous research the importance of those variables are new insights.

When looking at specific models and specific financial indicators it can be seen that features based on the Social Pillar Score are found to be important to predict the ROA in both Linear Regression and Random Forest. For this financial indicator Random Forest also found the 'ESG Score_Mean_10', 'Human Rights Score_Mean_3' and 'Human Rights Score_lag_3' to be important. In the case of the prediction of ROE both models found a variation of the Management Score to be important. Other than that Linear Regression found variations of the CSR Strategy Score to be important, as well as the 'Governance Pillar Score_diff'. While Random Forest found 'Human Rights Policy Score' and 'ESG Combined Score_Mean_3' to be important to predict ROE. Lastly, Linear Regression found 'Governance Pillar Score_Mean_10' to be important to predict the EPS, while Random Forest found 'Human Rights Score_Mean_10' to be important. All of these insights are obtained from the features importance graphs found in figs. 7.11 to 7.16.

6.2.2 Advanced Models

For the more advanced models it was found that the model with the 'ESG Combined Score' as exogenous variable improved predictions compared to the base model in almost every case for both ARIMA and LSTM. This was the benchmark scenario. Only in the case of the prediction of ROE with the ARIMA model the benchmark model did not improve. The prediction of the ROE in the ARIMA model improved for the Granger scenario, where the 'Resource Use Score_bin' variable was added as exogenous feature. In the case of the LSTM model the benchmark model improved for each of the financial indicators. Next to that the prediction of EPS improved in the case of the correlation and the Granger method, which included 'CO2 Equivalent Emissions Total' and 'CO2 Equivalent Emissions Total_Mean_10' respectively. The prediction of ROA improved when adding 'Human Rights Breaches Contractor' based on the correlation method. Lastly the prediction of ROE improved when adding 'Human Rights Policy Score' based on the union of the feature importance for both Linear Regression and Random Forest. Although the RMSE reduced for the previous mentioned scenarios, this result was not significant according to the permutation test, so no firm conclusions can be drawn.

6.3 Subquestion 3

Which financial performance can be predicted based on ESG performance?

6.3.1 Base Models

When looking at the RMSE for Linear Regression and Random Forest in table 5.1 if can be seen that for both models the prediction of the ROE and ROA outperforms the prediction of the EPS. Linear Regression makes its best prediction on the ROA, while Random Forest makes its best prediction on ROE. The RMSE of the prediction from Random Forest for ROA and ROE is however very similar. That is why on average the prediction of ROA is more accurate.

6.3.2 Advanced Models

It was found that in the case of ARIMA none of the financial indicators improved in more than one case. The prediction of the ROE improved in the case of the exogenous variable based on the Granger Causality test, while ROA and EPS improved in the benchmark scenario with the 'ESG Combined Score'. But once again, these improvements were not found to be significant.

In the case of LSTM for EPS there was improvement in every scenario except for the one based on the feature importance of the Linear Regression and Random Forest models. On the other hand the original models for this financial indicator were already performing poorly, so even though the new models are an improvement of the old models, they still performed inadequately.

The ROA financial indicator showed improvement in the LSTM model for 2 out of the 4 scenarios. Unlike the EPS model, the base model for ROA already performed well. Other then improvement in the benchmark models there was improvement found in the correlation based model. In this case the 'Human Rights Breaches Contractor' variable was included as exogenous variable for the ROA model. The ROE financial indicator also showed improvement in 2 out of 4 scenarios and, similar to ROA, performed well. For this financial indicator the improvement was in the benchmark model and the model based on the feature importance.

The prediction of EPS was consistently poor across all models and scenarios. Upon closer investigation is was found that the error in these cases mostly had to to with the error of the first fold. When looking at the graph for the final folds of EPS it can be seen that the fit for the AIRMA model is quite well. The highness of the overall error for this financial indicator for this model comes from the high error from the first fold.

On average the predictions of the ROA was found to be best for both models and all scenarios. ROE predictions followed in accuracy, while EPS predictions had the highest RMSE and were the least accurate. However in the case of the improvement of prediction when adding an exogenous variable, EPS improved in more cases.

In conclusion, there are a few financial indicators where the model including ESG performance outperforms the original model, but these results are not found to be significant.

6.4 Subquestion 4

Which selected method(s) works best to quantify the relationship between ESG and financial performance?

6.4.1 Base Models

When looking at the RMSE in figure 5.1 of the Linear Regression model and the Random Forest model, it can be seen that for each financial indicator Random Forest outperforms Linear Regression. This could indicate non linear and more complex relationship between ESG performance and financial performance.

6.4.2 Advanced Models

When looking at the average value of the RMSE of the ARIMA and LSTM models, it can be seen that overall ARIMA has a lower error. On the other hand in terms of the model including an exogenous variable outperforming the model without exogenous variable, LSTM performs better in most cases. These insights are obtained from table 5.4 and table 5.5.

For both models the prediction of EPS was relatively poor compared to the other financial indicators when looking at the average RMSE. However it can be seen that for ARIMA the prediction of the last fold seems very accurate when looking at the graphs of the last fold in figs. 7.17 to 7.20. This has to do with the prediction of the previous folds being inaccurate.

Next to that when looking at the last fold for ARIMA for the prediction of ROA, it can be seen that the prediction seems relatively bad, which can be found in figs. 7.33 to 7.36. In this case however the predictions of the other folds were better, which is why the average error of this scenario remains relatively low.

It can be concluded that ARIMA outperforms LSTM in terms of average RMSE in most cases, but LSTM outperforms ARIMA in terms of improvement in predictions when including an exogenous variable.

6.5 General Conclusion

This research contributes valuable insights into the complex relation between ESG performance and financial performance. With the use of Linear Regression and Random Forest the most important ESG features are determined for the models separately, but also the union of most important features was investigated. The union of these features can be found in 7.6. Here some expected features like the different pillar scores and general ESG Score can be found, but also some features providing new insights. These are for example features based on Human Rights, Product Responsibility, and Eco-Design. These are features considered important based on meetings with the KPMG sustainability team, which makes it an exciting outcome. Other features found to be important by Linear Regression and Random Forest are features based on the Management Score and CSR Strategy Score. The LSTM and ARIMA models are investigated to determine if the inclusion of ESG factors can improve the accuracy of the predictions of financial performance. These models found improvement in a few cases, but based on the permutation test these results were found to not be significant, so no firm conclusion can be drawn on this. Possible improvement of these models will be discussed in chapter 7, 'Discussion and Future Research'.

In conclusion, interesting ESG features were found to be important, which were already considered of great importance by KPMG, but not yet found to be of importance in previous research. Additionally, the predictive models might profit from the inclusion of ESG features, but these outcomes were found not significant yet. It would be valuable to perform more research on these features, and if these features end up being significant base ESG advice to KPMG's clients on these results.

7

Discussion and Future Research

This research investigated the relationship between the ESG performance and financial performance of companies. It was found that in some cases adding an exogenous ESG variables reduces the RMSE, but this result was not significant according to the Permutation Test. This chapter will discuss points of discussion and improvement, and will recommend next steps for future research.

7.1 Discussion

One of the limitations of the research was the data quantity. Even though data was available for the past 22 years, in the case of the ESG data this was only on a yearly base. This meant that the data was not very frequent, and not a lot of records were available. Especially for the LSTM model this was a disadvantage, since this method is known to need a lot of data to be able to perform well. This also meant that strong assumptions on the trend of the ESG data were made, which might not have been accurate to the actual behaviour of the data.

Another limitation of the research is the hyperparameter tuning of the LSTM model. For the tuning of the hyperparameters Grid Search was used. First of all the size of the grid was relatively small which made the combination of parameters limited. This had to do with the computational effort to apply grid search and the fact that this had to be done for multiple models and multiple financial indicators with limited time. Other than that grid search might not provide an optimal solution since the optimal solution might fall between the grid values and might not be found by the Grid Search method.

7.2 Future Research

The missing values in this research were handled by applying linear interpolation. This decision was based on simplicity of implementing and understanding, and due to lack of knowledge on the behaviour of the data. The research might have profited from applying another, more sophisticated imputation method to fill these missing values. Especially if it would be one that would have fitted the natural shape of the data better. For future research, for example the 'quadratic' or 'polynomial' method in the Pandas interpolate function could be tried, or other machine learning techniques to fit the data better.

Other than that, the way the most important features were selected was by using the Recursive Feature Elimination method for both Linear Regression and Random Forest. This method has as disadvantage that it is computationally intensive and sensitive to the initial model parameters. This problem can be addressed by trying different feature selection methods. For feature research alternative feature selection methods such as Lasso or Elastic Net could be tried.

Lastly, research was based on one company that KPMG work with. This means that the outcomes are not generalizable to all companies that KPMG work with. To be able to draw a more general conclusion on the relationship between ESG and financial performance, the performance of the models for different companies across different sectors should be considered. For future research it is interesting to look into more companies within different sectors. For this research only Microsoft was considered, which belongs to the 'Information Technology' sector. For companies in different sectors different ESG factors could be of importance. For future research it would be recommended to look into more different companies, to be able to draw a more general conclusion.

In conclusion, even though this research provided valuable insights into the relationship between ESG performance and financial performance, there are still limitations that would profit from a more extended research. The research could benefit from more frequent ESG data, improving feature selection and imputation methods, and adopting more advanced hyperparameter tuning techniques. In this way the robustness and generalizability of the models could be improved, resulting in better outcomes. Additionally, exploring more companies across different sectors could offer more general insights into the effect of ESG on financial performance.

References

[1] Tracy Dathe. *Implementing Environmental, Social and Governance (ESG) Principles for Sustainable Businesses: A Practical Guide in Sustainability Management.* Springer Nature, 2024. 1, 2

[2] Raisa Almeyda and Asep Darmansya. **The influence of environmental, social, and governance (ESG) disclosure on firm financial performance**. *IPTEK Journal of Proceedings Series*, (5):278–290, 2019. 1, 7, 8, 10, 11

[3] SFDR. **What is the SFDR?**, 2024. Accessed: 2024-05-23. 1

[4] CSRD. **Corporate Sustainability Reporting Directive**, 2024. Accessed: 2024-05-23. 1

[5] Gunnar Friede, Timo Busch, and Alexander Bassen. **ESG and financial performance: aggregated evidence from more than 2000 empirical studies**. *Journal of sustainable finance & investment*, **5**(4):210–233, 2015. 2, 10

[6] Changhong Zhao, Yu Guo, Jiahai Yuan, Mengya Wu, Daiyu Li, Yiou Zhou, and Jiangang Kang. **ESG and corporate financial performance: Empirical evidence from China's listed power generation companies**. *Sustainability*, **10**(8):2607, 2018. 2, 10

[7] Daniel C Esty and Todd Cort. *Values at work: Sustainable investing and ESG reporting.* Springer, 2020. 2, 8

[8] Ting-Ting Li, Kai Wang, Toshiyuki Sueyoshi, and Derek D Wang. **ESG: Research progress and future prospects**. *Sustainability*, **13**(21):11663, 2021. 7

[9] EBA. **EBA report on management and supervision of ESG risks for credit institutions and investment firms**, 2021. 7

[10] Dixon Cooper Mitchell Franklin (CPA), Patty Graybeal. *Principles of Accounting, Volume 1.* OpenStax, Rice University, 2023. 8, 9

[11] Dyah Purnamasari. **The effect of changes in return on assets, return on equity, and economic value added to the stock price changes and its impact on earnings per share**. *Research journal of finance and accounting*, **6**(6):80–90, 2015. 8, 9

[12] Agha Ammad Nabi. **Earnings per share impact on non-financial firms Performance**. *Journal of Economic Info*, **1**(4):1–7, 2014. 9

[13] Balamuralikrishnan Chakkravarthy, Francis Gnanasekar Irudayasamy, Rajesh Elangovan, Natarajan Rengaraju, and Satyanarayana Parayitam. **Relationship between return on assets and firm value: Institutional holdings and firm size as moderators**. *Quality & Quantity*, pages 1–17, 2023. 10

[14] Tensie Whelan, Ulrich Atz, Tracy Van Holt, and Casey Clark. **ESG and financial performance**. *Uncovering the Relationship by Aggregating Evidence from*, **1**:2015–2020, 2021. 10

[15] LSEG. **Environmental, Social and Governance Scores**, 2024. Accessed: 2024-03-01. 15, 16

[16] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. **LOF: identifying density-based local outliers**. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 93–104, 2000. 17

[17] Alice Zheng and Amanda Casari. *Feature engineering for machine learning: principles and techniques for data scientists.* " O'Reilly Media, Inc.", 2018. 18, 19, 20

[18] Mark Hoogendoorn and Burkhardt Funk. **Machine learning for the quantified self**. *On the art of learning from sensory data*, 2018. 18

[19] Zijian Zhang, Jinzhong Yang, Angela Ho, Wen Jiang, Jennifer Logan, Xin Wang, Paul D Brown, Susan L McGovern, Nandita Guha-Thakurta, Sherise D Ferguson, et al. **A predictive model for distinguishing radiation necrosis from tumour progression after gamma knife radiosurgery based on radiomic features from MR images**. *European radiology*, **28**:2255–2263, 2018. 19

[20] Andrew Gelman. **Scaling regression inputs by dividing by two standard deviations**. *Statistics in medicine*, **27**(15):2865–2873, 2008. 20

[21] Mihaela Rosca, Theophane Weber, Arthur Gretton, and Shakir Mohamed. **A case for new neural network smoothness constraints**. 2020. 20

[22] Sven F Crone and Nikolaos Kourentzes. **Feature selection for time series prediction–A combined filter and wrapper approach for neural networks**. *Neurocomputing*, **73**(10-12):1923–1936, 2010. 21

[23] Baptiste Gregorutti, Bertrand Michel, and Philippe Saint-Pierre. **Correlation and variable importance in random forests**. *Statistics and Computing*, **27**:659–678, 2017. 21

[24] Pablo M Granitto, Cesare Furlanello, Franco Biasioli, and Flavia Gasperi. **Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products**. *Chemometrics and intelligent laboratory systems*, **83**(2):83–90, 2006. 21

[25] David J Olive and David J Olive. *Multiple linear regression*. Springer, 2017. 23, 24

[26] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, **2**. Springer, 2009. 23, 25

[27] Peter Martin. **Linear regression: An introduction to statistical models**. 2022. 24, 25

[28] Gérard Biau and Erwan Scornet. **A random forest guided tour**. *Test*, **25**:197–227, 2016. 25

[29] Leo Breiman. **Random forests**. *Machine learning*, **45**:5–32, 2001. 26

[30] Wiwik Anggraeni, Retno Aulia Vinarti, and Yuni Dwi Kurniawati. **Performance comparisons between arima and arimax method in moslem kids clothes demand forecasting: Case study**. *Procedia Computer Science*, **72**:630–637, 2015. 26, 27, 28

[31] Ruey S Tsay. *Analysis of financial time series*. John wiley & sons, 2005. 27

[32] Garima Jain and Bhawna Mallick. **A study of time series models ARIMA and ETS**. *Available at SSRN 2898968*, 2017. 28

[33] Williams Kwasi Adu, Peter Appiahene, and Stephen Afrifa. **VAR, ARIMAX and ARIMA models for nowcasting unemployment rate in Ghana using Google trends**. *Journal of Electrical Systems and Information Technology*, **10**(1):12, 2023. 28

[34] Bruce H Andrews, Matthew D Dean, Robert Swain, and Caroline Cole. **Building ARIMA and ARIMAX models for predicting long-term disability benefit application rates in the public/private sectors**. *Society of Actuaries*, pages 1–54, 2013. 28

[35] Shu-Ling Lin and Xiao Jin. **Does ESG predict systemic banking crises? A computational economics model of early warning systems with interpretable multi-variable LSTM based on mixture attention**. *Mathematics*, **11**(2):410, 2023. 29, 30

[36] Filippo Maria Bianchi, Enrico Maiorino, Michael C Kampffmeyer, Antonello Rizzi, and Robert Jenssen. **Recurrent neural networks for short-term load forecasting: an overview and comparative analysis**. 2017. 29, 30

[37] Ruijie Huang, Chenji Wei, Baohua Wang, Jian Yang, Xin Xu, Suwei Wu, and Suqi Huang. **Well performance prediction based on Long Short-Term Memory (LSTM) neural network**. *Journal of Petroleum Science and Engineering*, **208**:109686, 2022. 29, 31

[38] Sepp Hochreiter and Jürgen Schmidhuber. **Long short-term memory**. *Neural computation*, **9**(8):1735–1780, 1997. 30

[39] ZhiQiang Li, HongXia Zou, and Bin Qi. **Application of ARIMA and LSTM in relative humidity prediction**. In *2019 IEEE 19th International Conference on Communication Technology (ICCT)*, pages 1544–1549. IEEE, 2019. 30, 31

[40] Soumik Ray and Banjul Bhattacharyya. **Statistical modeling and forecasting of ARIMA and ARIMAX models for food grains production and net availability of India.** 2020. 31, 32

[41] Matskevichus Mariia and Gladilin Peter. **Model Selection Approach for Time Series Forecasting**. In *2019 IEEE 13th International Conference on Application of Information and Communication Technologies (AICT)*, pages 1–5. IEEE, 2019. 31

[42] He-Ren Lou, Xin Wang, Ya Gao, and Qiang Zeng. **Comparison of ARIMA model, DNN model and LSTM model in predicting disease burden of occupational pneumoconiosis in Tianjin, China**. *BMC Public Health*, **22**(1):2167, 2022. 31, 32

[43] Aidan Meyler, Geoff Kenny, and Terry Quinn. **Forecasting Irish inflation using ARIMA models**. 1998. 31

[44] Chainarong Amornbunchornvej, Elena Zheleva, and Tanya Berger-Wolf. **Variable-lag granger causality and transfer entropy for time series analysis**. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, **15**(4):1–30, 2021. 32

[45] John R Freeman. **Granger causality and the times series analysis of political relationships**. *American Journal of Political Science*, pages 327–358, 1983. 32

[46] Statsmodels. **statsmodels.tsa.stattools.grangercausalitytests**, 2024. Accessed: 2024-06-06. 32

[47] Philipp Probst, Marvin N Wright, and Anne-Laure Boulesteix. **Hyperparameters and tuning strategies for random forest**. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, **9**(3):e1301, 2019. 32, 33

[48] Tadayoshi Fushiki. **Estimation of prediction error by using K-fold cross-validation**. *Statistics and Computing*, **21**:137–146, 2011. 33

[49] Ron Kohavi et al. **A study of cross-validation and bootstrap for accuracy estimation and model selection**. In *Ijcai*, **14**, pages 1137–1145. Montreal, Canada, 1995. 33

[50] Hoang Lan Vu, Kelvin Tsun Wai Ng, Amy Richter, and Chunjiang An. **Analysis of input set characteristics and variances on k-fold cross validation for a Recurrent Neural Network model on waste disposal rate estimation**. *Journal of environmental management*, **311**:114869, 2022. 33

[51] Scikit-learn. **TimeSeriesSplit**, 2024. Accessed: 2024-06-28. 33

[52] Kenneth J Berry, Janis E Johnston, and Paul W Mielke. *A primer of permutation statistical methods.* Springer, 2019. 34, 39

Appendix

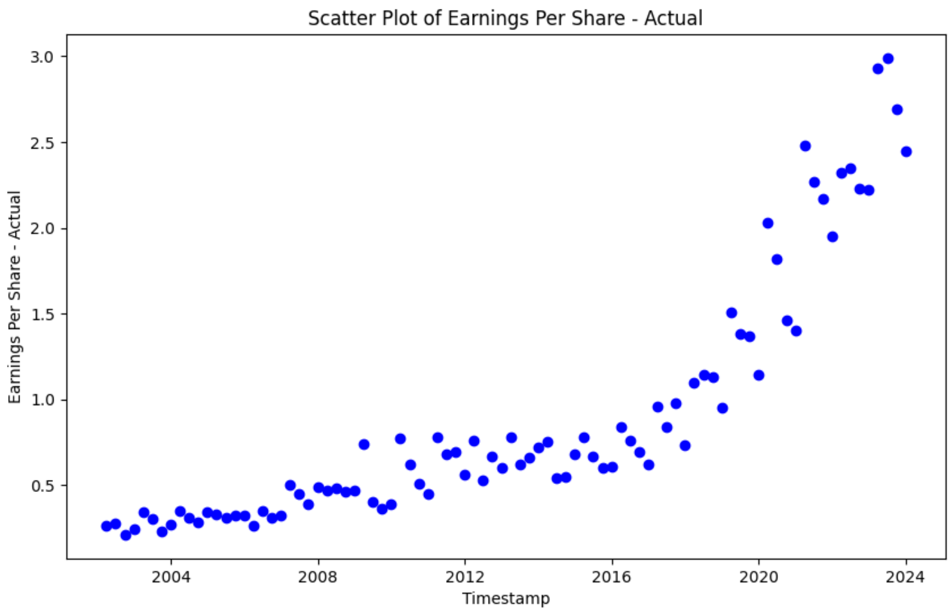

Figure 7.1: Scatterplot of Earning per Share

Name	Description
ESG Score	Refinitiv ESG Score is an overall company score based on the self-reported information in the environmental, social and corporate governance pillars.
ESG Controversies Score	ESG controversies category score measures a company's exposure to environmental, social and governance controversies and negative events reflected in global media.
ESG Combined Score	Refinitiv ESG Combined Score is an overall company score based on the reported information in the environmental, social and corporate governance pillars (ESG Score) with an ESG Controversies overlay.
Social Pillar Score	The social pillar measures a company's capacity to generate trust and loyalty with its workforce, customers and society, through its use of best management practices. It is a reflection of the company's reputation and the health of its license to operate, which are key factors in determining its ability to generate long term shareholder value.
Governance Pillar Score	The corporate governance pillar measures a company's systems and processes, which ensure that its board members and executives act in the best interests of its long term shareholders. It reflects a company's capacity, through its use of best management practices, to direct and control its rights and responsibilities through the creation of incentives, as well as checks and balances in order to generate long term shareholder value.
Environmental Pillar Score	The environmental pillar measures a company's impact on living and non-living natural systems, including the air, land and water, as well as complete ecosystems. It reflects how well a company uses best management practices to avoid environmental risks and capitalize on environmental opportunities in order to generate long term shareholder value.
Emissions Score	Emission category score measures a company's commitment and effectiveness towards reducing environmental emission in the production and operational processes.
Resource Use Score	Resource use category score reflects a company's performance and capacity to reduce the use of materials, energy or water, and to find more eco-efficient solutions by improving supply chain management.
Environmental Innovation Score	Environmental innovation category score reflects a company's capacity to reduce the environmental costs and burdens for its customers, and thereby creating new market opportunities through new environmental technologies and processes or eco-designed products.
Workforce Score	Workforce category score measures a company's effectiveness towards job satisfaction, healthy and safe workplace, maintaining diversity and equal opportunities, and development opportunities for its workforce.
Human Rights Score	Human rights category score measures a company's effectiveness towards respecting the fundamental human rights conventions.
Community Score	Community category score measures the company's commitment towards being a good citizen, protecting public health and respecting business ethics.
Product Responsibility Score	Product responsibility category score reflects a company's capacity to produce quality goods and services integrating the customer's health and safety, integrity and data privacy.
Management Score	Management category score measures a company's commitment and effectiveness towards following best practice corporate governance principles.
Shareholders Score	Shareholders category score measures a company's effectiveness towards equal treatment of shareholders and the use of anti-takeover devices.
CSR Strategy Score	CSR strategy category score reflects a company's practices to communicate that it integrates the economic (financial), social and environmental dimensions into its day-to-day decision-making processes.

Table 7.1: General ESG Scores Overview

Name	Description
CO2 Equivalent Emissions Total	Total Carbon dioxide (CO_2) and CO_2 equivalents emission in tonnes. • Total CO_2 emission = direct (scope1) + indirect (scope 2) • We follow greenhouse gas (GHG) protocol for all our emission classifications by type
Policy Human Rights	Does the company have a policy to ensure the respect of human rights in general? • Information to be on ensuring the respect of human rights • Consider a process on general fundamental human rights
Policy Human Rights Score	Does the company have a policy to ensure the respect of human rights in general? • Information to be on ensuring the respect of human rights • Consider a process on general fundamental human rights
Human Rights Policy	Does the company have a policy for the exclusion of child, forced or compulsory labour, or to guarantee the freedom of association universally applied independent of local laws?
Human Rights Contractor	Does the company report or show to use human rights criteria in the selection or monitoring process of its suppliers or sourcing partners? • Information to be on using human rights criteria while selecting a supplier or sourcing materials from sourcing partners. • Consider information from industry code such as the Electronic Industry Citizenship Coalition (EICC) code of conduct and Pharmaceutical Industry Principles (PSCI)
Human Rights Breaches Contractor	Does the company report or show to be ready to end a partnership with a sourcing partner if human rights criteria are not met? • Information to be on termination of contract or partnership with the supplier when human rights criteria are not met • Consider information from industry code such as the Electronic Industry Citizenship Coalition (EICC) code of conduct and Pharmaceutical Industry Principles (PSCI)
Human Rights Controversies	Number of controversies published in the media linked to human rights issues. • Controversies of suppliers and contractors relating to human rights breach • Human rights abuses/violations are considered
Human Rights Policy Score	Does the company have a policy for the exclusion of child, forced or compulsory labour, or to guarantee the freedom of association universally applied independent of local laws?
Eco-Design Products	Does the company report on specific products which are designed for reuse, recycling or the reduction of environmental impacts? • Products that have been specifically designed with the goal of being recycled, reused or which are disposed of without negatively impacting the environment • There must be some discussion of environmental concerns during the product design
Biodiversity Impact Reduction	Does the company report on its impact on biodiversity or on activities to reduce its impact on the native ecosystems and species, as well as the biodiversity of protected and sensitive areas?

Table 7.2: Additional ESG Scores Overview

Column	Missing Values
ESG Score	1
ESG Combined Score	1
ESG Controversies Score	1
Social Pillar Score	1
Governance Pillar Score	1
Environmental Pillar Score	1
Emissions Score	1
Resource Use Score	1
Environmental Innovation Score	1
Workforce Score	1
Human Rights Score	1
Community Score	1
Product Responsibility Score	1
Management Score	1
Shareholders Score	1
CSR Strategy Score	1
CO2 Equivalent Emissions Total	4
Policy Human Rights	1
Policy Human Rights Score	1
Human Rights Policy	1
Human Rights Contractor	1
Human Rights Breaches Contractor	1
Human Rights Controversies	17
Human Rights Policy Score	1
Eco-Design Products	1
Biodiversity Impact Reduction	1

Table 7.3: Missing values count per column for ESG data.

Column	Missing Values
Return On Equity - Actual	2
Return On Assets - Actual	9
Earnings Per Share - Actual	0

Table 7.4: Missing values count per column for financial data.

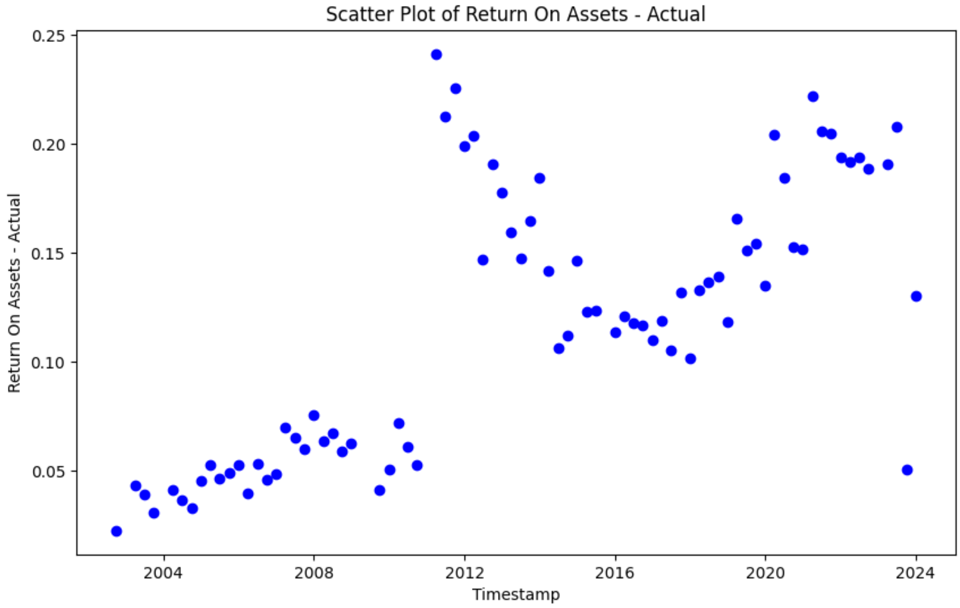

Figure 7.2: Scatterplot of Return on Assets

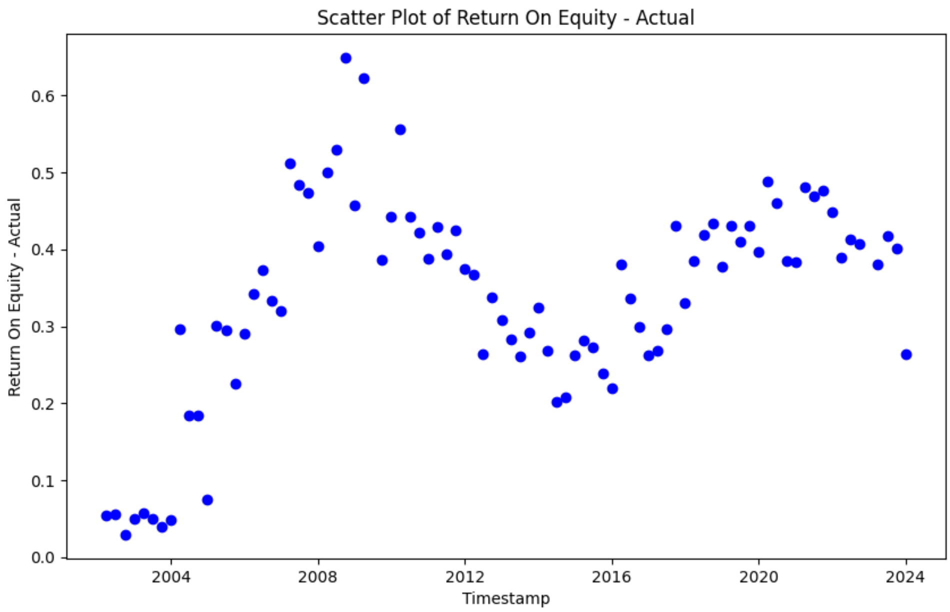

Figure 7.3: Scatterplot of Return on Equity

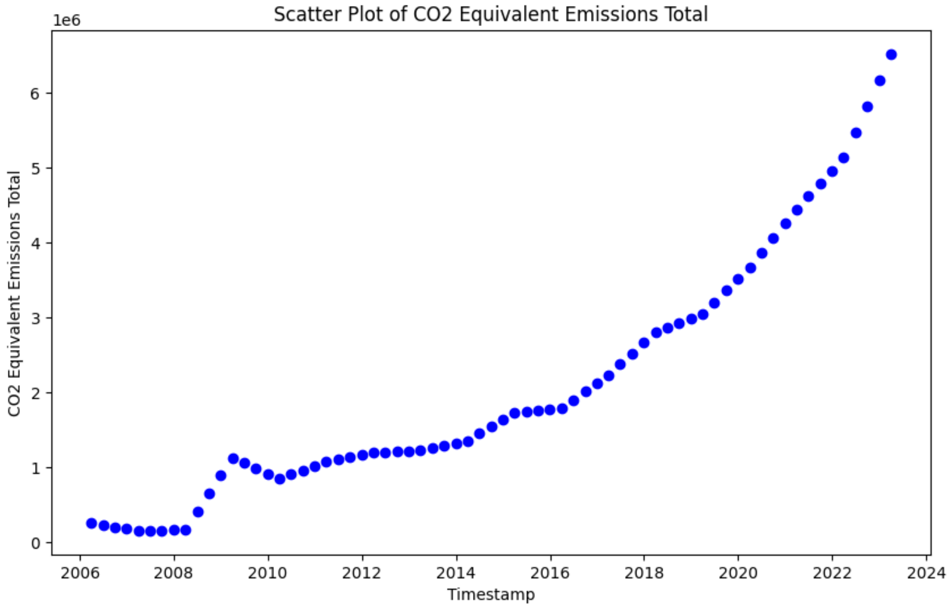

Figure 7.4: Scatterplot of CO2 Equivalent Emissions Total

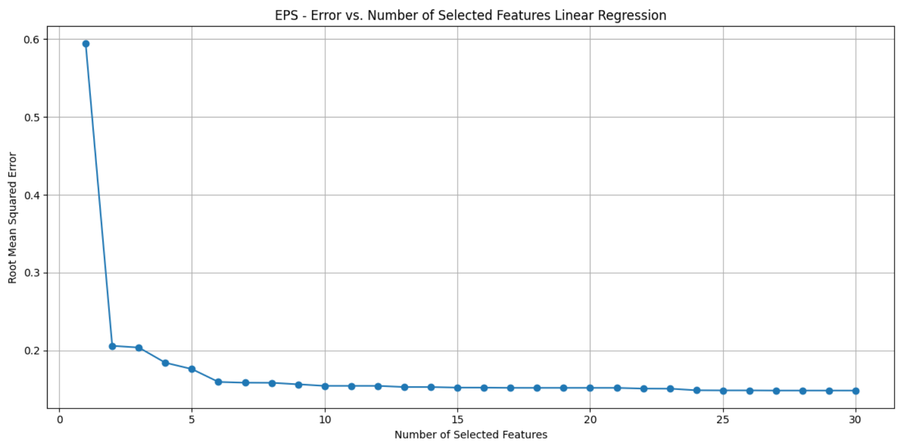

Figure 7.5: Errors for Linear Regression model - EPS

Name	Description
column_Mean	Mean value for column calculated over 2 periods
column_Mean_3	Mean value for column calculated over 3 periods
column_Mean_10	Mean value for column calculated over 10 periods
column_Percentage	Percentage change value for column
column_diff	Absolute difference in value for column
column_diff_cat	Categorized difference in value for column
column_lag_1	Lagged value for column (1 period)
column_lag_2	Lagged value for column (2 periods)
column_lag_3	Lagged value for column (3 periods)
column_bin	Binary value for column
Environmental_Social_average	Average value for Environmental and Social factors
Environmental_Governance_average	Average value for Environmental and Governance factors
Social_Governance_average	Average value for Social and Governance factors
Human_Rights_Policy_interaction	Interaction between column and Human Rights Policy
Human_Rights_Controversies_interaction	Interaction between column and Human Rights Controversies
Eco_Design_Policy_interaction	Interaction between column and Eco Design Policy
Bio_Diversity_Policy_interaction	Interaction between column and Biodiversity Policy
CO2_Environmental_Interaction	Interaction between column and CO2 Environmental factors

Table 7.5: Newly engineered features

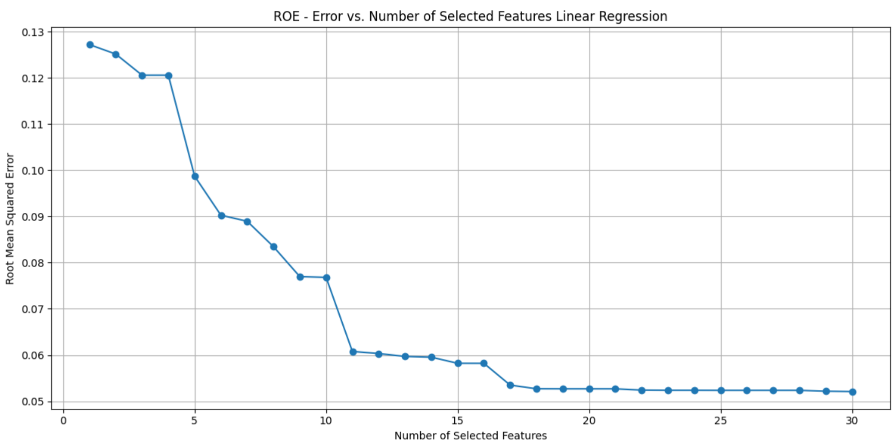

Figure 7.6: Errors for Linear Regression model - ROE

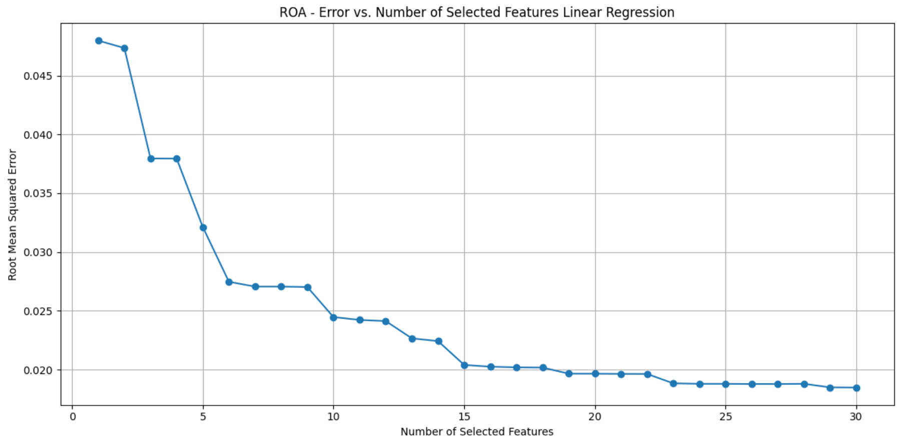

Figure 7.7: Errors for Linear Regression model - ROA

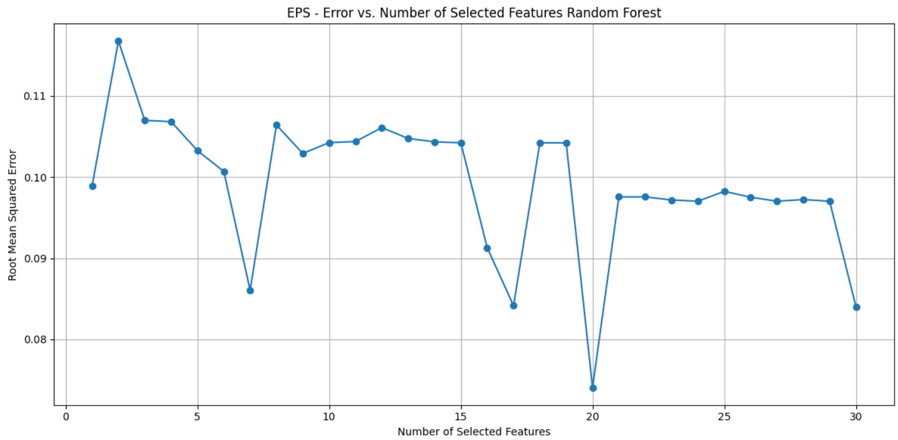

Figure 7.8: Errors for Random Forest model - EPS

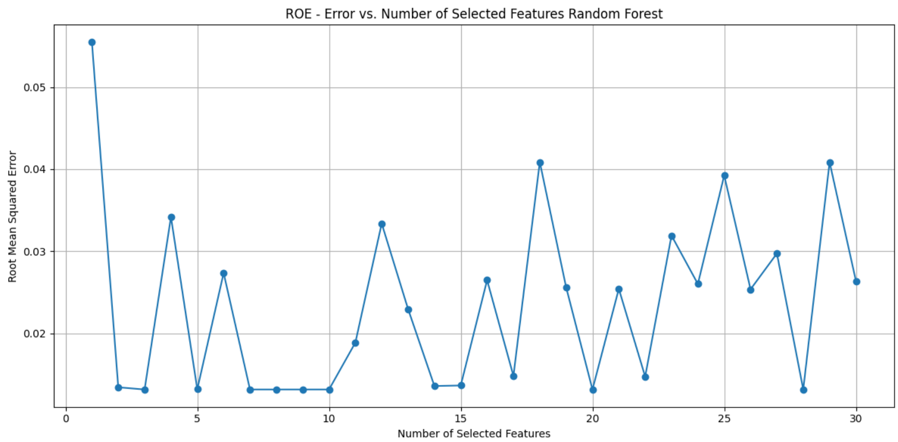

Figure 7.9: Errors for Random Forest model - ROE

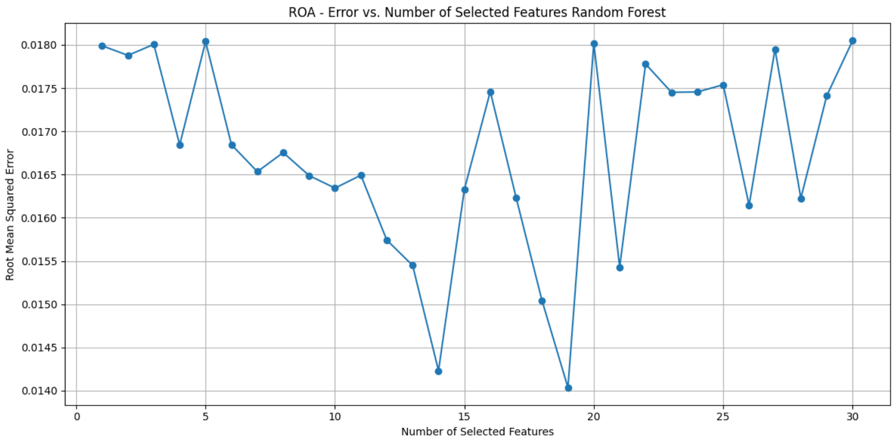

Figure 7.10: Errors for Random Forest model - ROA

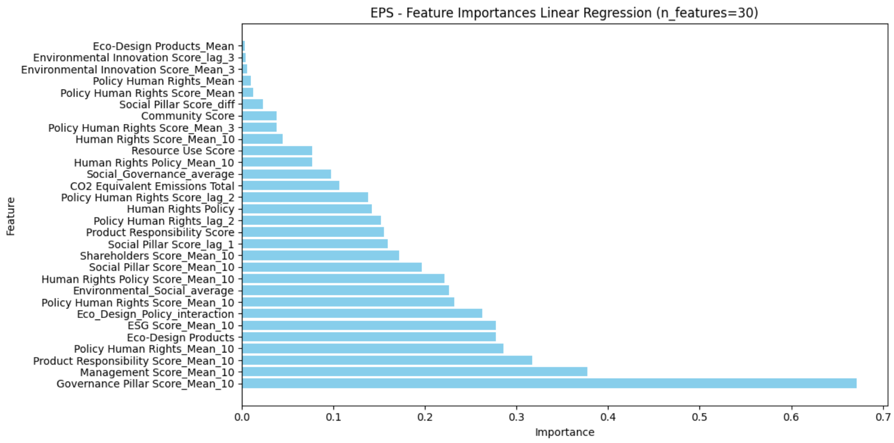

Figure 7.11: Feature importance of Linear Regression - EPS

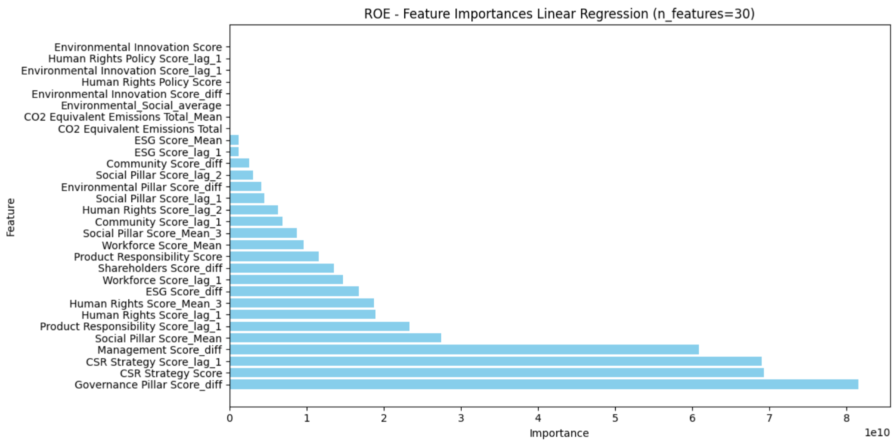

Figure 7.12: Feature importance of Linear Regression - ROE

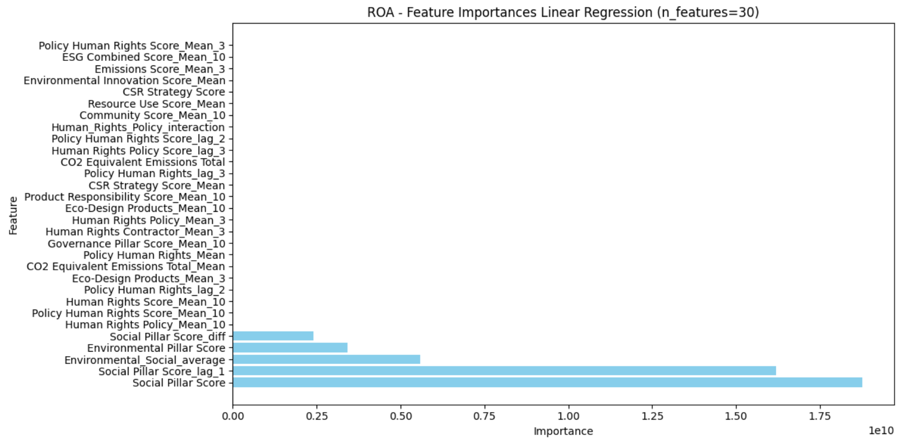

Figure 7.13: Feature importance of Linear Regression - ROA

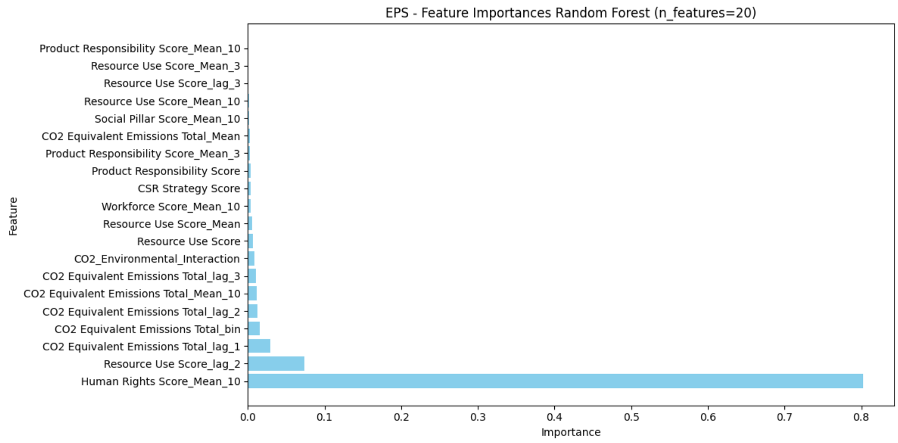

Figure 7.14: Feature importance of Random Forest - EPS

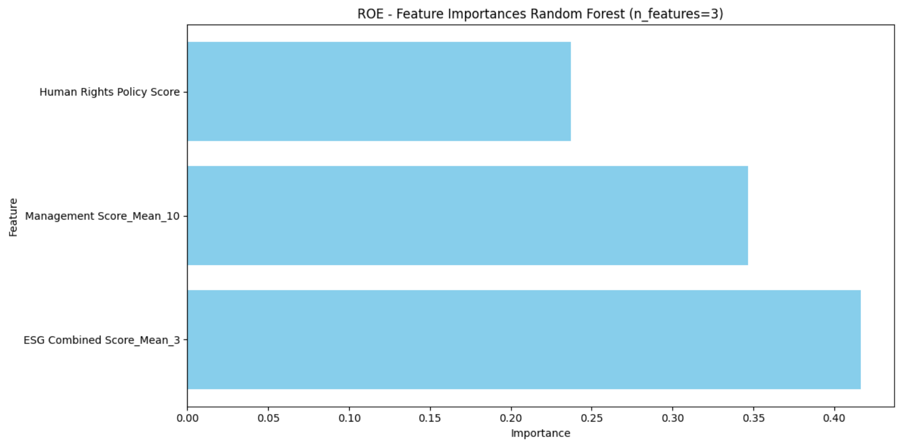

Figure 7.15: Feature importance of Random Forest - ROE

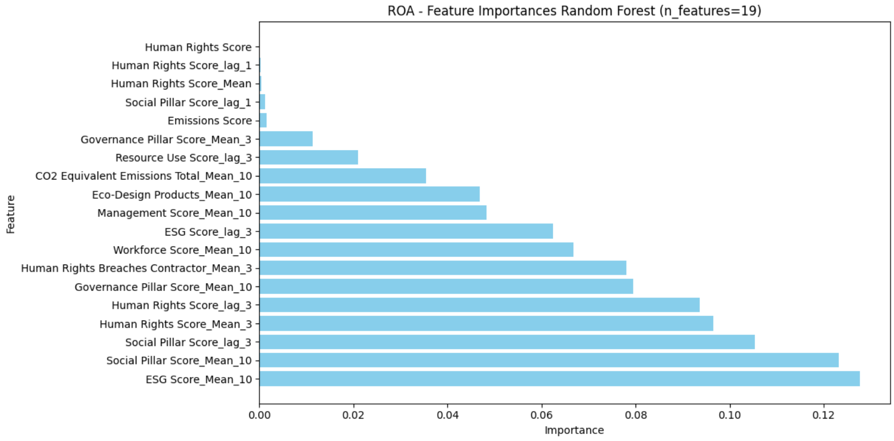

Figure 7.16: Feature importance of Random Forest - ROA

Indicator	**Exogenous Variables**
EPS	Human Rights Score_Mean_10, Resource Use Score, Product Responsibility Score, Social Pillar Score_Mean_10, Product Responsibility Score_Mean_10
ROE	Human Rights Policy Score
ROA	Eco-Design Products_Mean_10, Social Pillar Score_lag_1, Governance Pillar Score_Mean_10

Table 7.6: Union of best found ESG features

EPS	**ROE**	**ROA**
CO2 Equivalent Emissions Total_Mean_10	Resource Use Score_bin	Human Rights Breaches Contractor_Mean_3
CO2_Environmental_Interaction	Environmental Pillar Score_lag_3	Human Rights Breaches Contractor_lag_1
CO2 Equivalent Emissions Total_bin	Resource Use Score_Mean_10	Human Rights Breaches Contractor_Mean
CO2 Equivalent Emissions Total	Product Responsibility Score_lag_3	Eco-Design Products_lag_3
CO2 Equivalent Emissions Total_Mean_3	Environmental Innovation Score_lag_3	Human Rights Breaches Contractor

Table 7.7: Best exogenous variables based on the Granger Causality test

EPS	**ROE**	**ROA**
CO2 Equivalent Emissions Total	Emissions Score_Mean_3	Human Rights Breaches Contractor
CO2 Equivalent Emissions Total_bin	Emissions Score_bin	Human Rights Breaches Contractor_Mean
CO2 Equivalent Emissions Total_Mean	Emissions Score_lag_1	Eco-Design Products_lag_3
CO2 Equivalent Emissions Total_Mean_3	Emissions Score_Mean	Human Rights Breaches Contractor_lag_1
CO2 Equivalent Emissions Total_lag_1	Emissions Score	Human Rights Breaches Contractor_Mean_3

Table 7.8: Best exogenous variables based on the correlation matrix

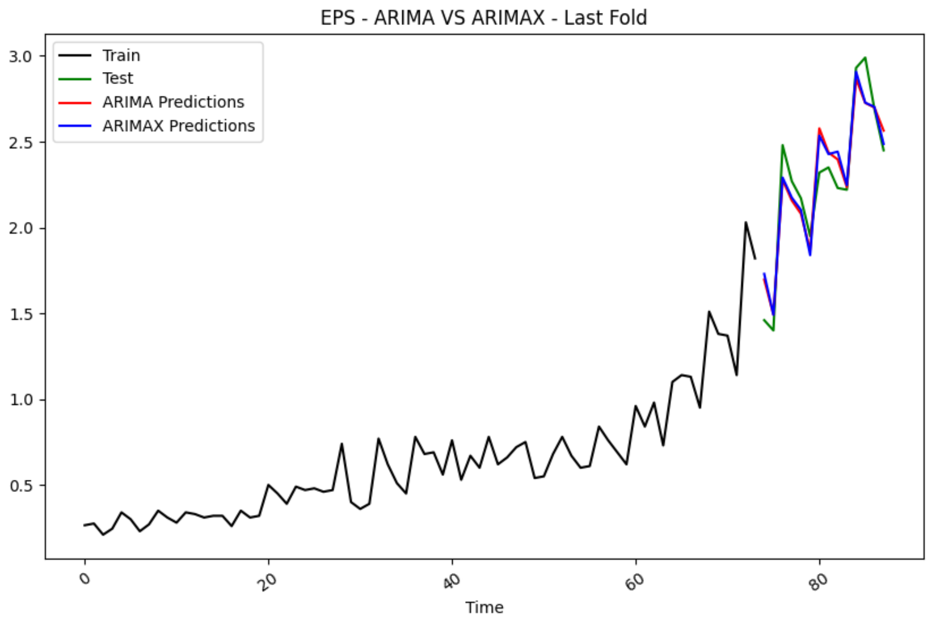

Figure 7.17: EPS ARIMA VS ARIMAX (exogenous variable based on correlation) last fold

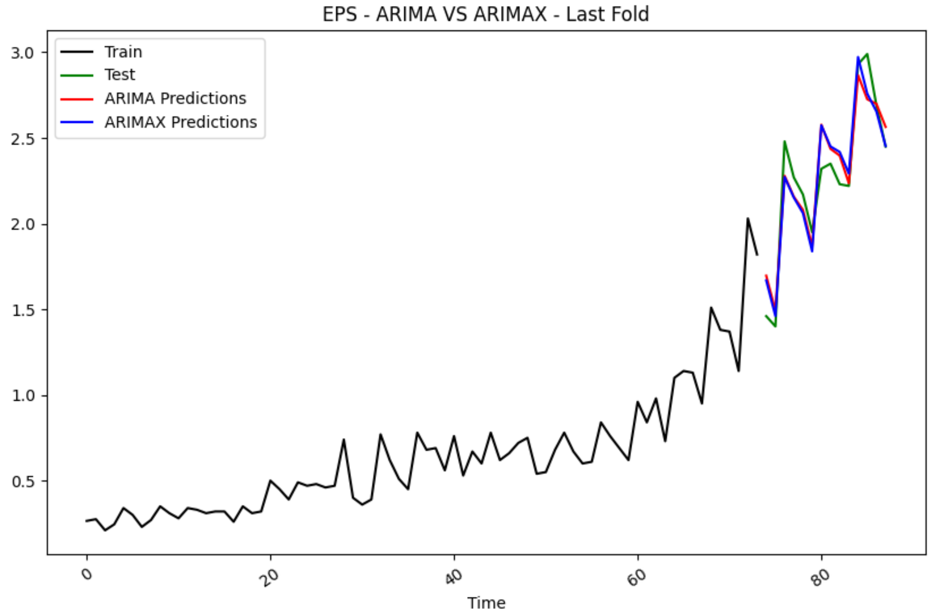

Figure 7.18: EPS ARIMA VS ARIMAX (exogenous variable based on Granger Causality test) last fold

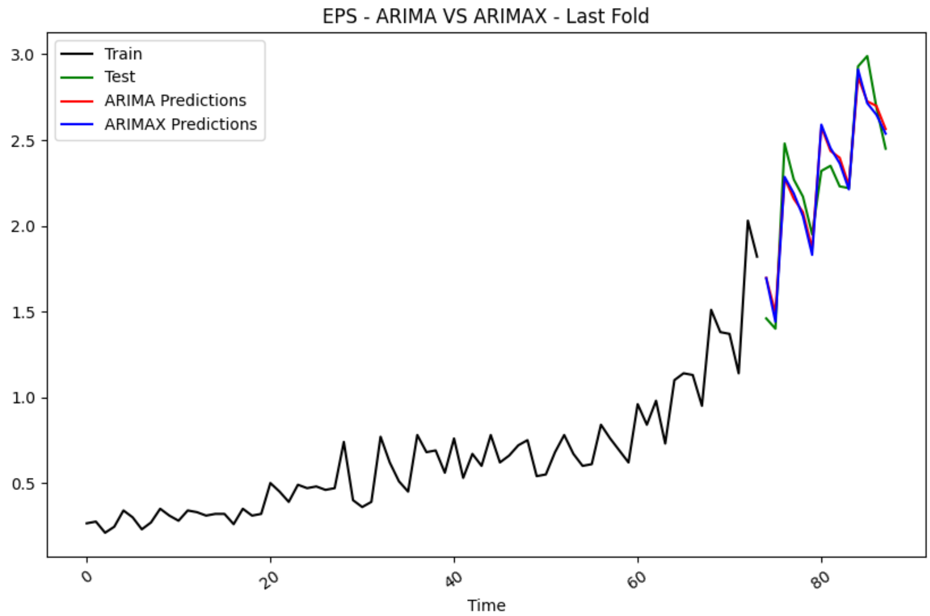

Figure 7.19: EPS ARIMA VS ARIMAX (exogenous variable based on Base Model) last fold

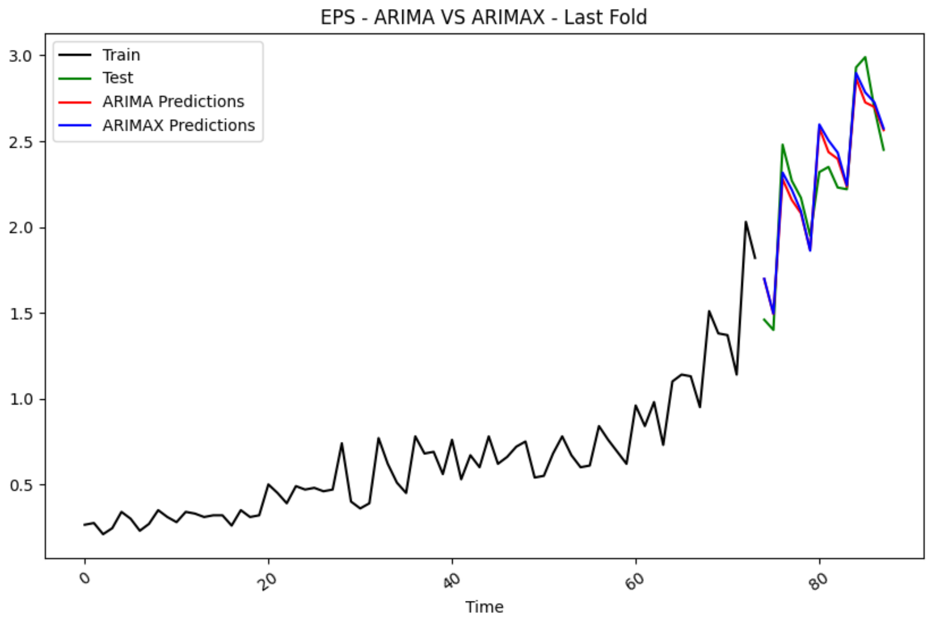

Figure 7.20: EPS ARIMA VS ARIMAX (exogenous variable based on 'ESG Combined Score') last fold

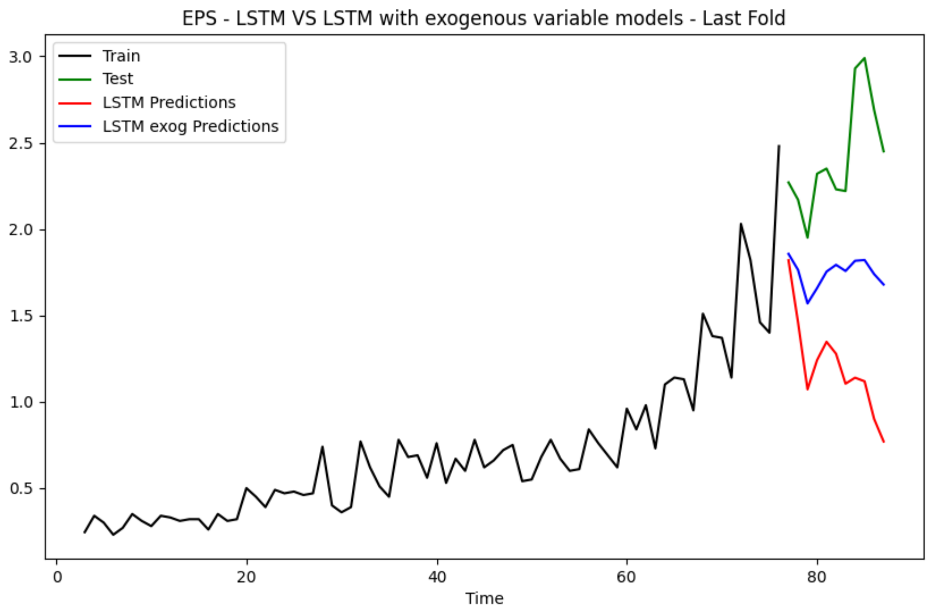

Figure 7.21: EPS LSTM VS LSTMX (exogenous variable based on correlation) last fold

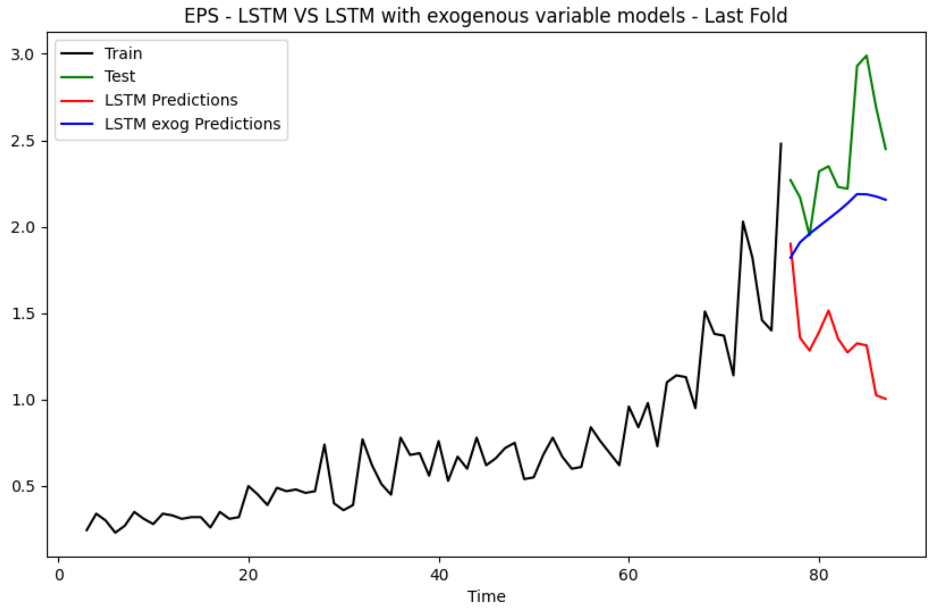

Figure 7.22: EPS LSTM VS LSTMX (exogenous variable based on Granger Causality test) last fold

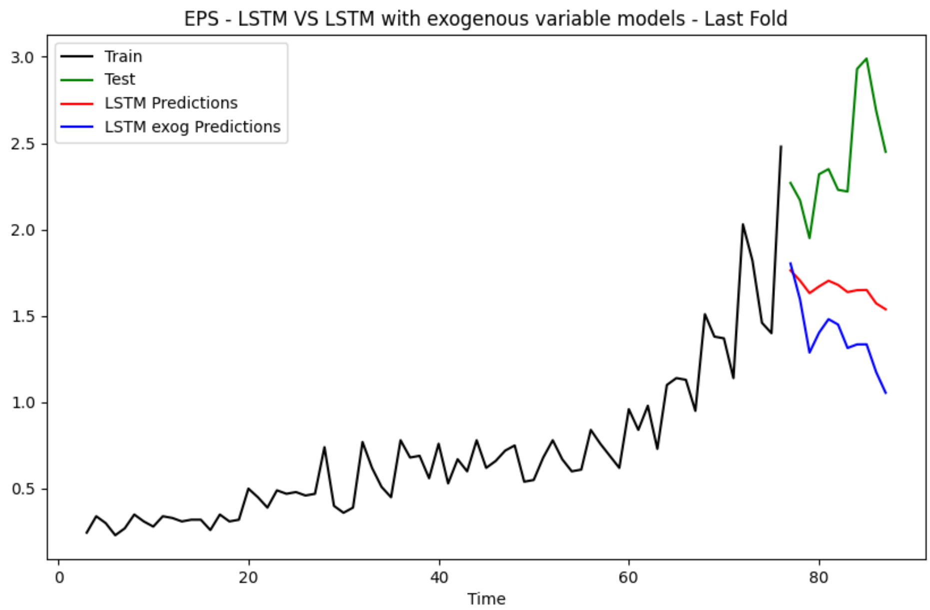

Figure 7.23: EPS LSTM VS LSTMX (exogenous variable based on Base Model) last fold

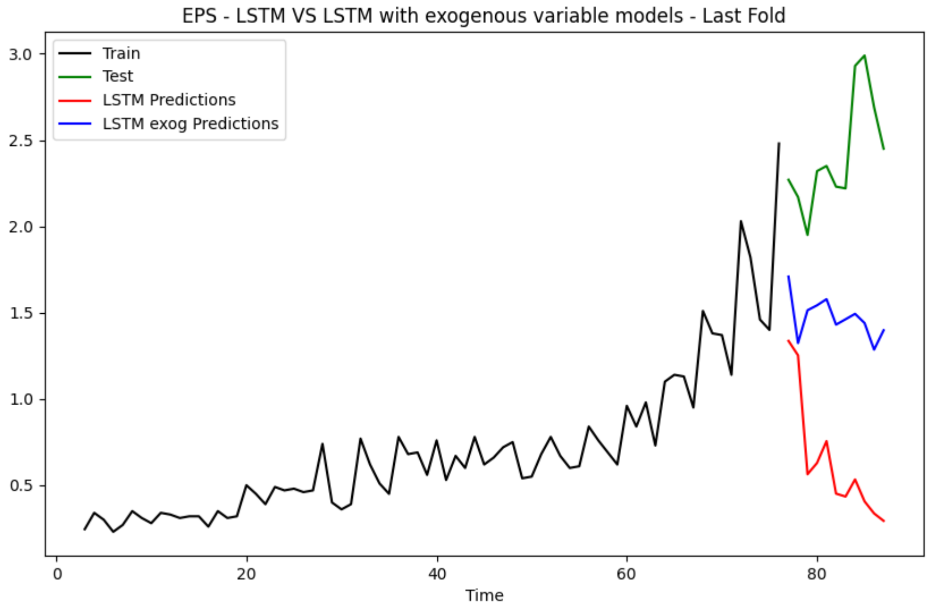

Figure 7.24: EPS LSTM VS LSTMX (exogenous variable based on 'ESG Combined Score') last fold

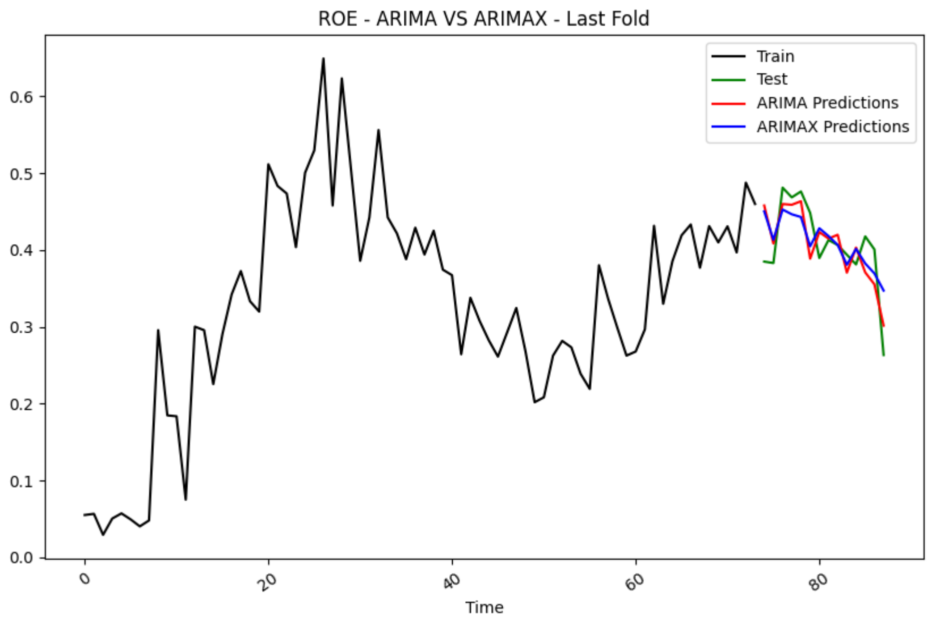

Figure 7.25: ROE ARIMA VS ARIMAX (exogenous variable based on correlation) last fold

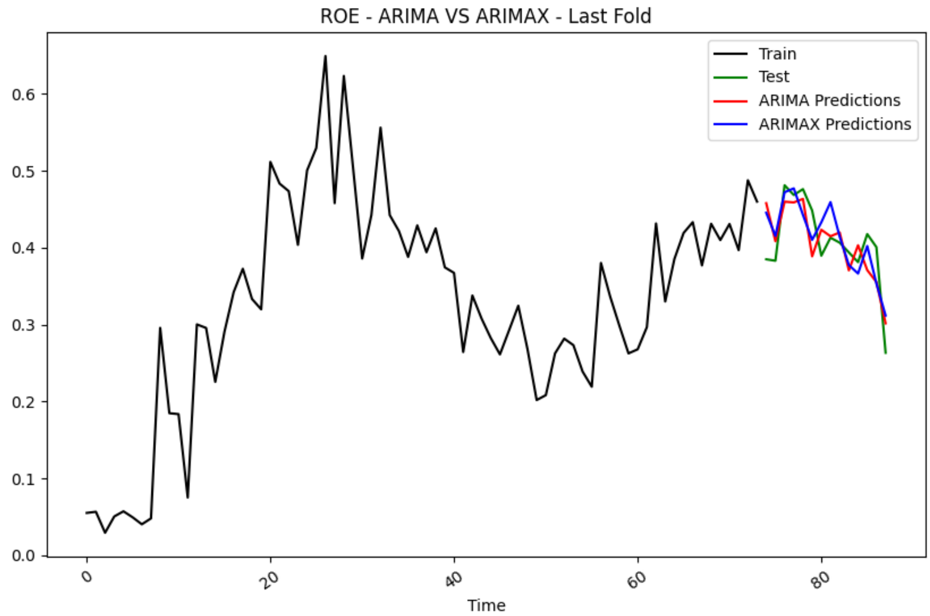

Figure 7.26: ROE ARIMA VS ARIMAX (exogenous variable based on Granger Causality test) last fold

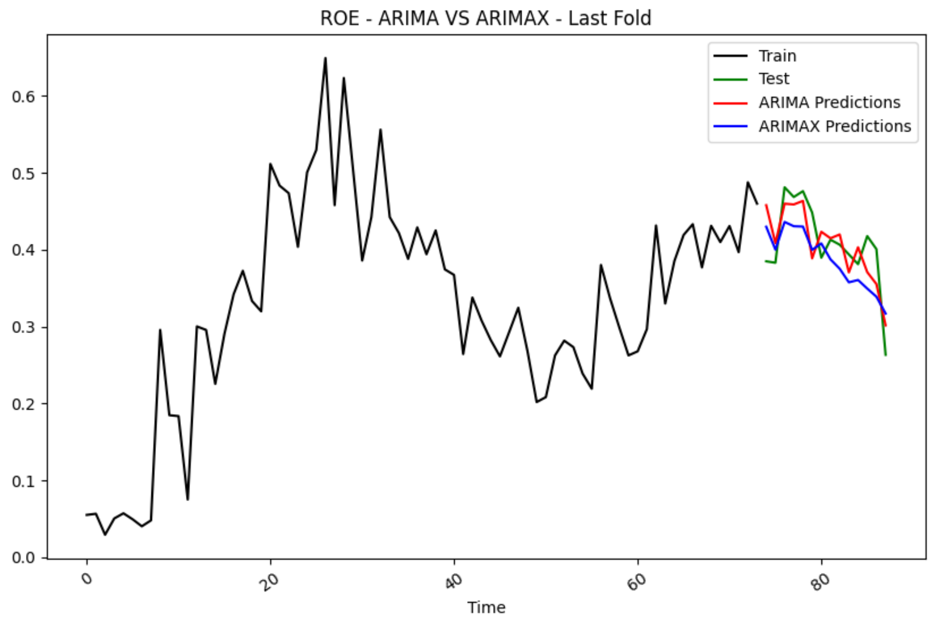

Figure 7.27: ROE ARIMA VS ARIMAX (exogenous variable based on Base Models) last fold

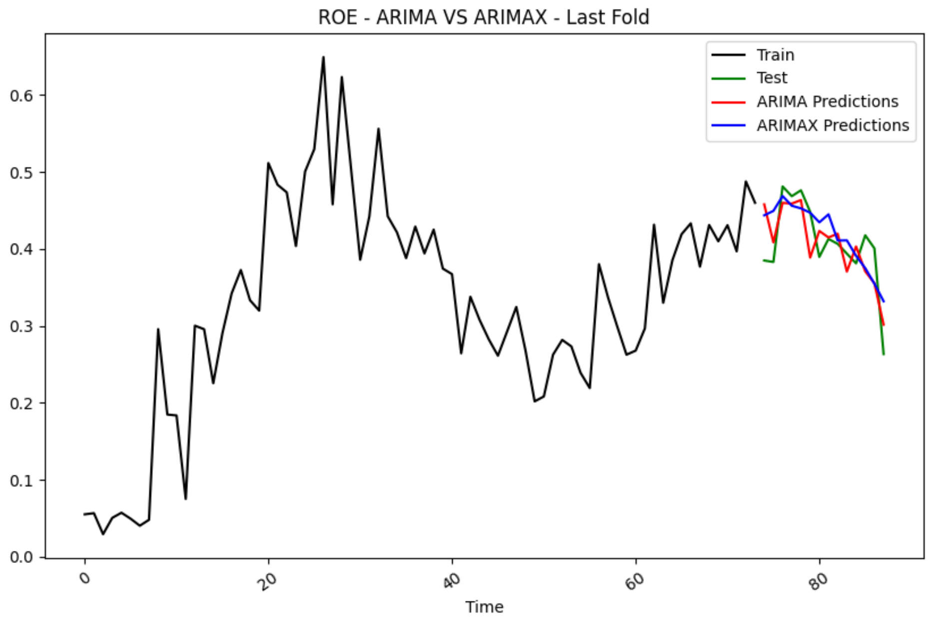

Figure 7.28: ROE ARIMA VS ARIMAX (exogenous variable based on 'ESG Combined Score') last fold

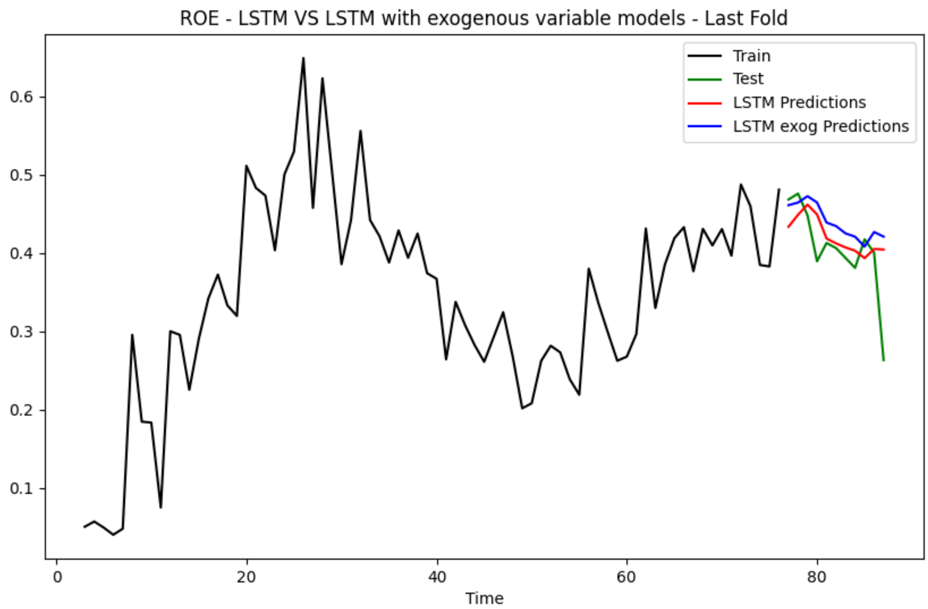

Figure 7.29: ROE LSTM VS LSTMX (exogenous variable based on correlation) last fold

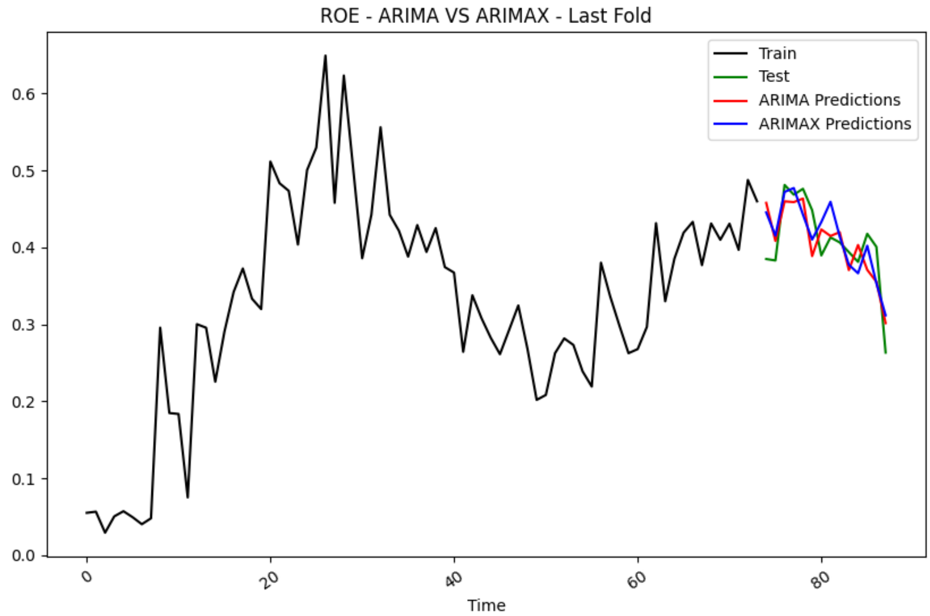

Figure 7.30: ROE LSTM VS LSTMX (exogenous variable based on Granger Causality test) last fold

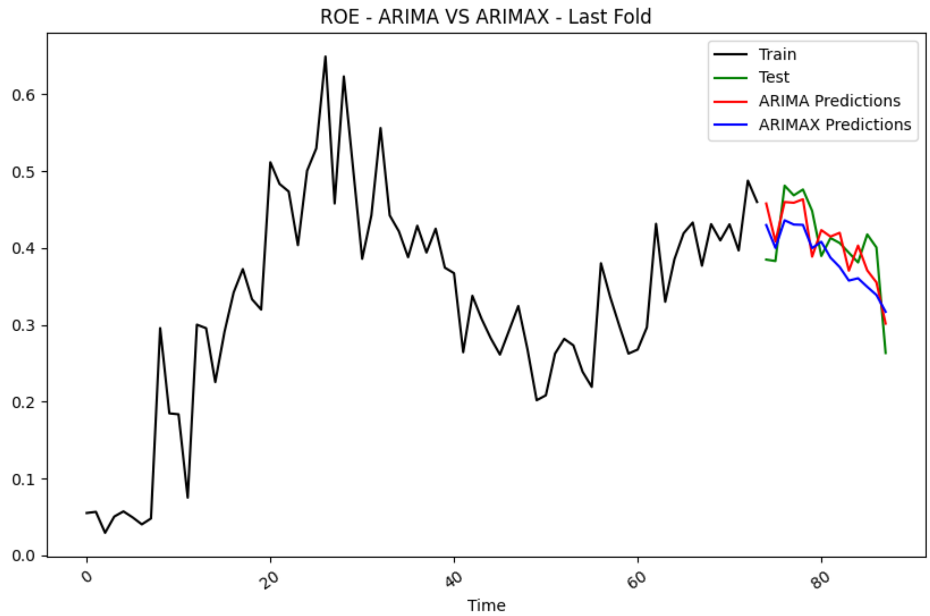

Figure 7.31: ROE LSTM VS LSTMX (exogenous variable based on Base Models) last fold

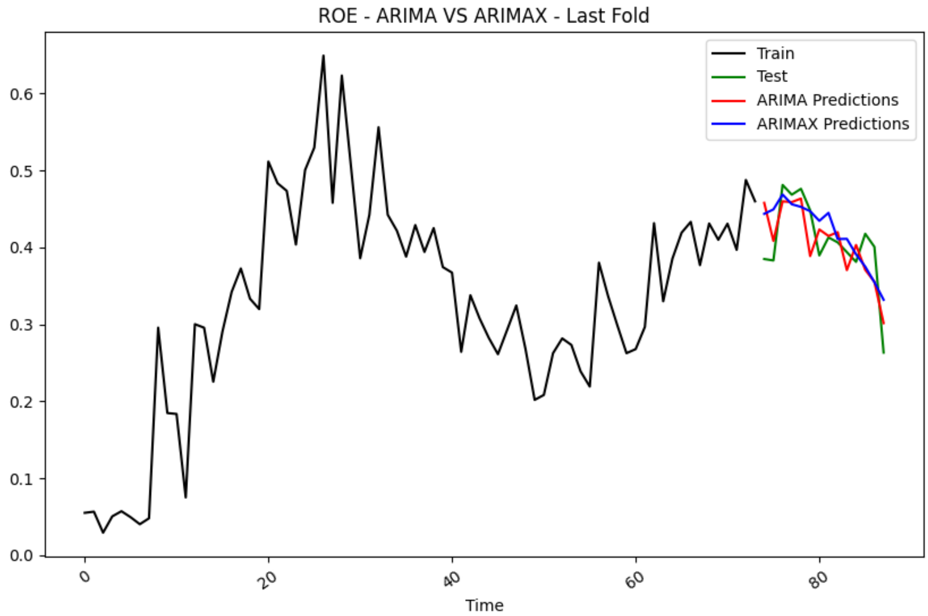

Figure 7.32: ROE LSTM VS LSTMX (exogenous variable based on 'ESG Combined Score') last fold

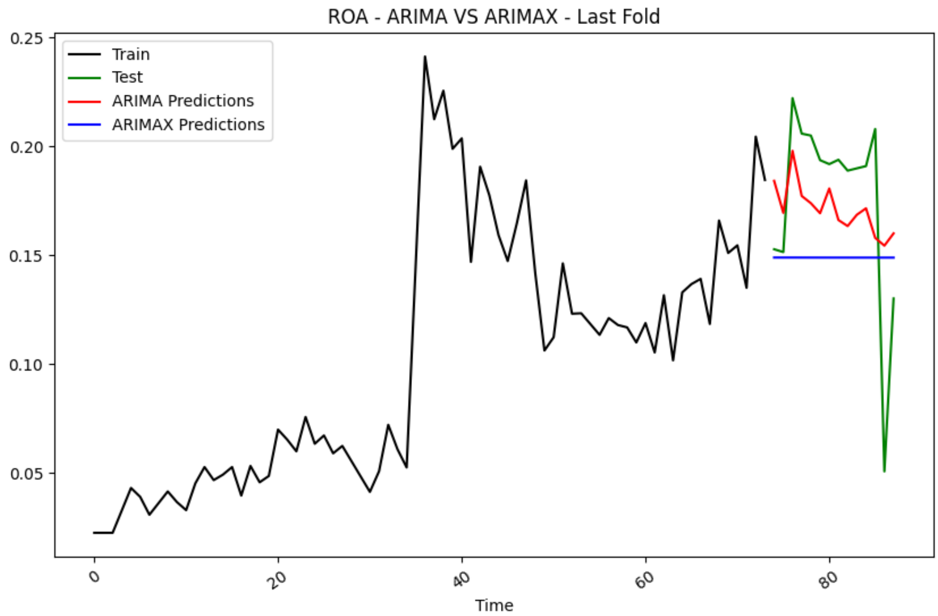

Figure 7.33: ROA ARIMA VS ARIMAX (exogenous variable based on correlation) last fold

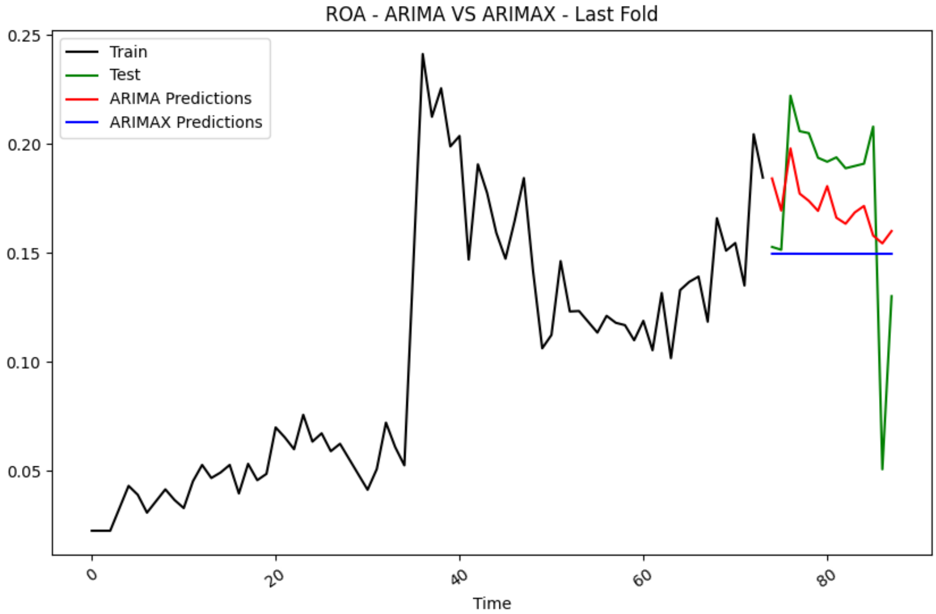

Figure 7.34: ROA ARIMA VS ARIMAX (exogenous variable based on Granger Causality test) last fold

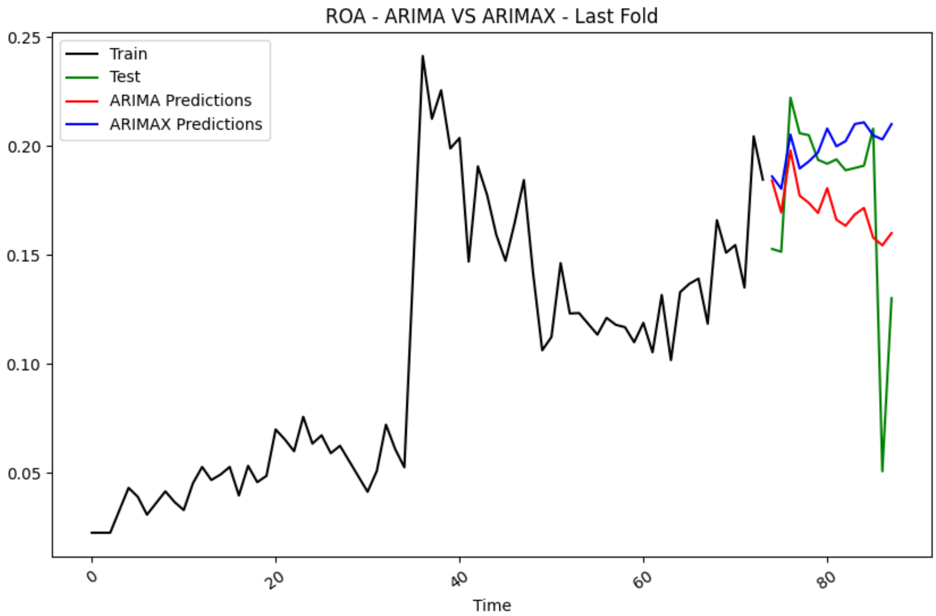

Figure 7.35: ROA ARIMA VS ARIMAX (exogenous variable based on Base Models) last fold

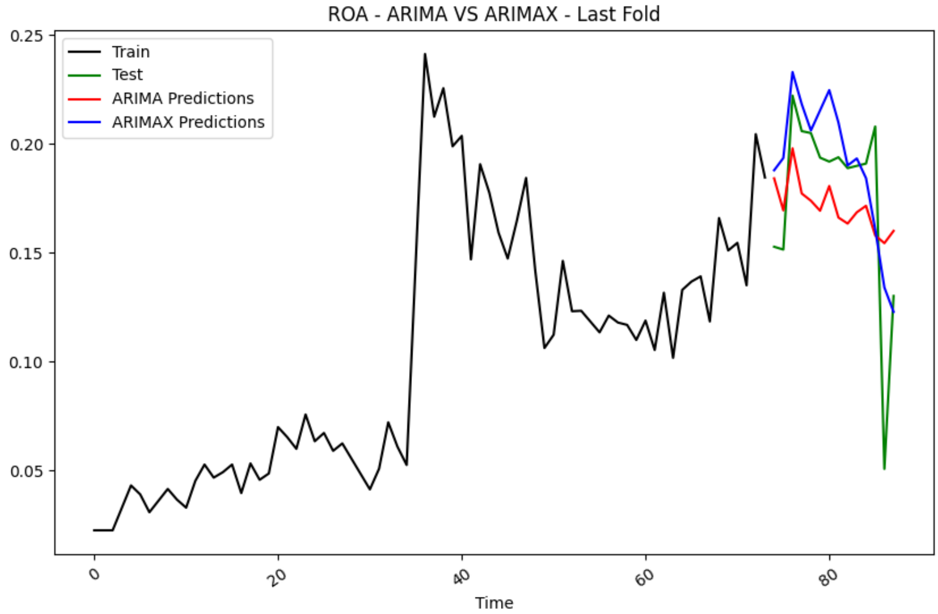

Figure 7.36: ROA ARIMA VS ARIMAX (exogenous variable based on 'ESG Combined Score') last fold

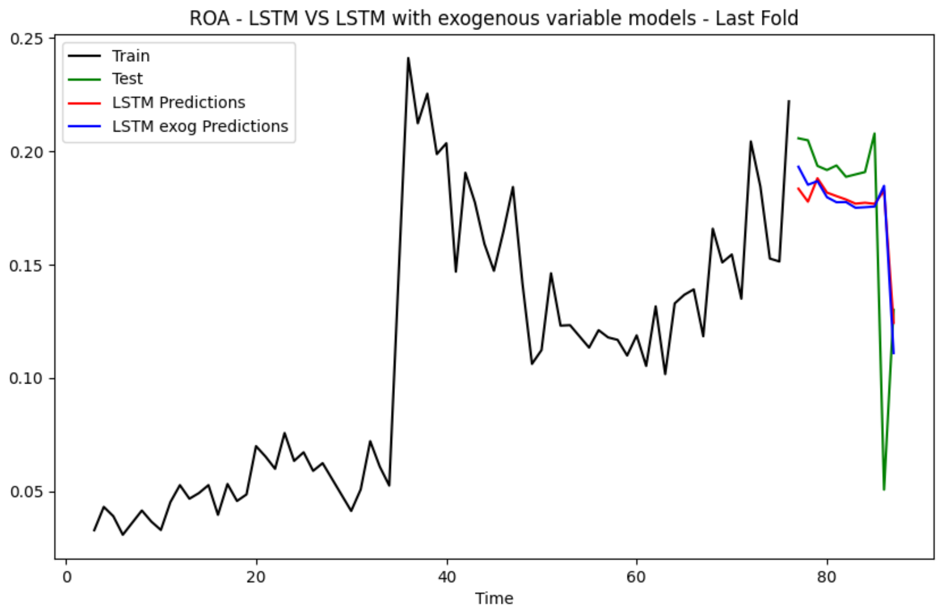

Figure 7.37: ROA LSTM VS LSTMX (exogenous variable based on correlation) last fold

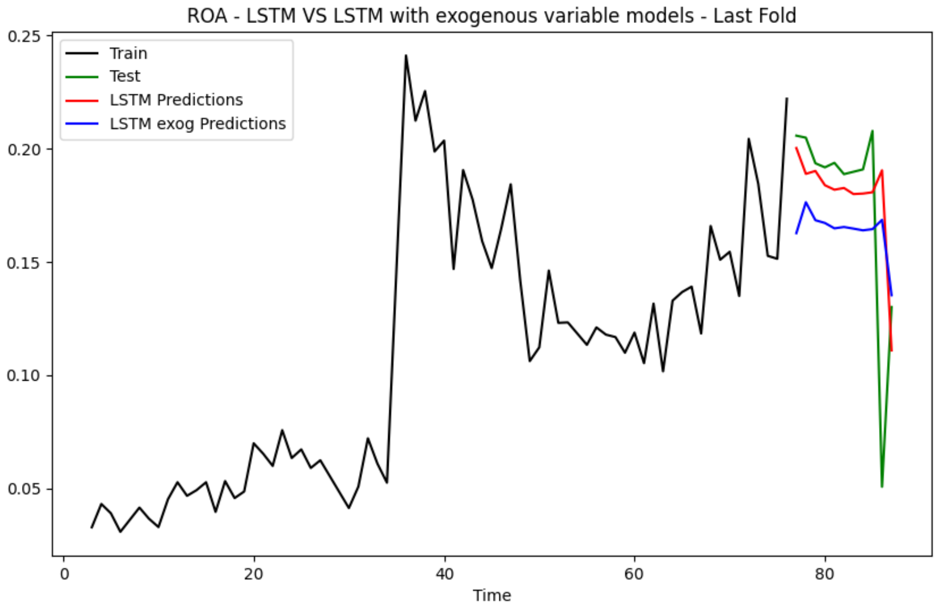

Figure 7.38: ROA LSTM VS LSTMX (exogenous variable based on Granger Causality test) last fold

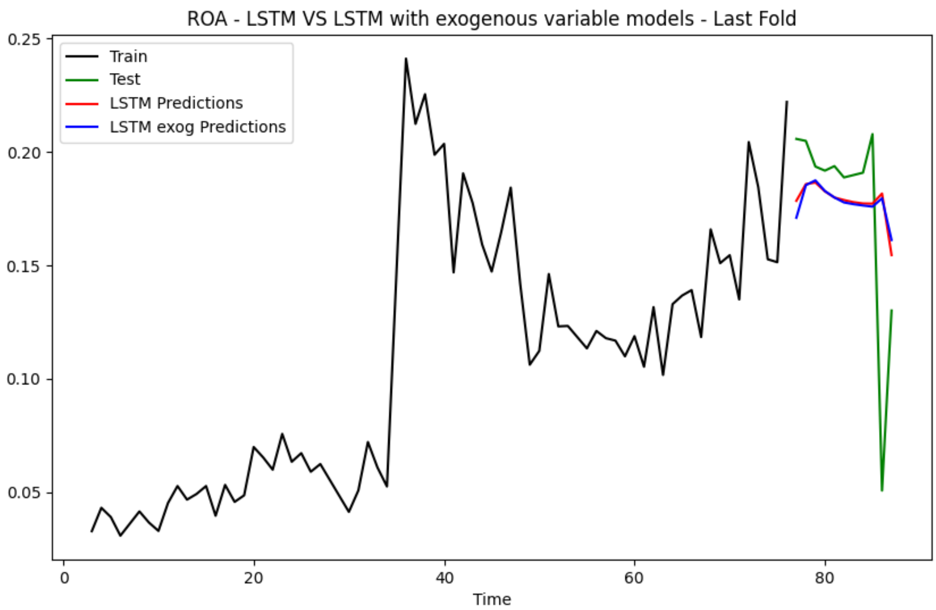

Figure 7.39: ROA LSTM VS LSTMX (exogenous variable based on Base Models) last fold

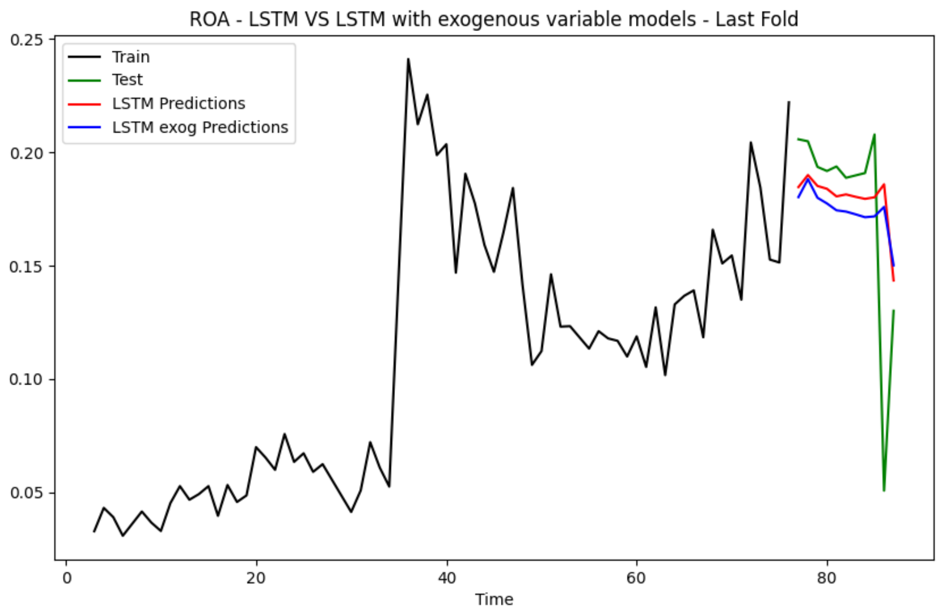

Figure 7.40: ROA LSTM VS LSTMX (exogenous variable based on 'ESG Combined Score') last fold