Erasmus University Rotterdam
Erasmus School of Economics

The hidden costs of the corporate carbon footprint: a machine learning approach

Master Thesis

MSc Econometrics and Management Science
Specialization: Quantitative Finance

Author:
Julius A. T. van Bebber
508138

Supervisor:
Prof. Dr. D.J.C. van Dijk

Second Assessor:
dr. R. Lange

October 5, 2021

Abstract

Companies are more pressured than ever to reduce their greenhouse gas (GHG) emissions by investors and regulators. There has been a twentyfold increase in the number of climate change (related) laws since 2000 and investors are including the score on Environment, Social and Governance of a company in their investment decisions. However, not all companies disclose their GHG emissions. This study uses machine learning to create a model to forecast the corporate carbon footprint across regions, industries and sectors. Light Gradient Boosting Machine showed the best prediction performance using multiple publicly available predictor variables to estimate GHG emissions. Compared to existing linear models, the mean absolute error is reduced by up to 13%. Next, the hidden costs of the corporate carbon footprint as valued by investors are looked into. This study uses the corporate carbon footprint as a predictor to explore the contemporaneous relation with equity value. Afterwards, coefficients, SHapley Additive exPlanation (SHAP) values and first-order derivatives are used to explore predictor relations. Using linear regression, a negative price elasticity of up to -0.053% is found between the carbon footprint and equity value. This negative relation is also indicated by the SHAP values. Further analysis using first-order derivatives showed a non-linear relation indicating the existence of a threshold. Emissions above this threshold have a much smaller negative or even a slightly positive impact on equity value indicating a smaller equity value discount for heavy polluting companies.

Keywords: corporate carbon footprint, equity value discount, machine learning, Light Gradient Boosting Machine, neural networks

Contents

1 Introduction

Climate change is seen as the top global threat facing our planet at this moment (Poushter and Huang (2019)). Of all the companies disclosing to the Carbon Disclosure Project in 2019 53% identified climate-related risks (CDP (2019)). As increasing concentrations of greenhouse gases in the atmosphere are causing climate change, carbon accounting has become a popular topic in industrial ecology research. However, the methods for corporate carbon footprint calculations are still evolving. According to Pandey et al. (2011), there is no standard method used in practice, resulting in little coherence in definitions and calculations of carbon footprints. Furthermore, they state that corporates include different gases in their footprint calculations and that the disclosed scopes of emissions vary making reports incomparable.

This study focuses on two aspects of the corporate carbon footprint. First, this study forecasts the corporate carbon footprint of companies that do not disclose their emissions using machine learning. By examining predictor importance, variables are identified that have a large impact on the corporate carbon footprint. Second, this paper will look into the hidden costs of the corporate carbon footprint as valued by investors. It will use the corporate carbon footprint as a predictor to explore the contemporaneous relation with equity value. In this way, the potential hidden costs of carbon emission as valued by investors are shown using the former relation. Several methods are analyzed in order to find the models with the highest prediction performance. The two models that will be estimated will make use of publicly available data across regions, industries and sectors. Having a unified estimation method is of importance considering the growing regulatory demands from regulators around the world. Next to regulators, there is more pressure from investors, society and corporates too to disclose the corporate carbon footprint (Giese et al. (2019)).

On the one hand, growing attention worldwide comes with potential negative future risks. The risks, identified by 53% of the CEOs disclosing to the Carbon Disclosure Project in 2019, can be categorised into two categories: transition risk and physical risk. Physical risk follows from environmental changes due to climate change such as extreme weather and rising global temperature. For most companies, transition risk focuses on the potential policy and legal changes. For instance, transition risk could include future emission reporting obligations showing the importance of having a unified and proven estimation method. On the other hand, companies identified opportunities as well. The potential benefits even outweigh the potential costs of the negative risks since companies are able to offer new products and services, operate on new markets and can offer low emission

products and services.

Not only companies acknowledge the potential risks and opportunities, so do investors. Several papers have been written on the relation between the corporate carbon footprint and equity value. They find a negative relation for U.S. Firms (Matsumura et al. (2014)), European firms (Clarkson et al. (2013)) and Australian firms (Chapple et al. (2013)). The second model in this study will explore this relation too. However, it uses data across regions, industries and sectors eliminating the sector and region-specific limitations of the previous studies. It is also the first study that uses machine learning methods to explore this relation.

Not every company discloses its corporate carbon footprint and this can have several reasons. Heavy polluting companies can choose not to disclose because of liability exposure or competitive disadvantage. Another reason can be that the company does not have the resources available to calculate its corporate carbon footprint or it is only able to calculate its direct emissions. The World Research Institute and World Business Council for Sustainable Development (2004) identify these direct emissions as scope 1 emissions: direct emissions from company facilities and company vehicles. Scope 2 emissions are indirect emissions from purchased electricity, steam, heating and cooling for own use. Lastly, scope 3 emissions are indirect emissions from up- and downstream activities such as purchased goods, investments and employee commuting.

Previous studies include different gases under the name corporate carbon footprint. In this study, the definition of the corporate carbon footprint as defined in Wiedmann and Minx (2008) is used. They define the corporate carbon footprint as a measure of the exclusive total amount of carbon dioxide emissions that are directly and indirectly caused by the activities or products from corporates. They only focus on carbon dioxide since many of the other greenhouse gas emissions are not based on carbon or are harder to quantify due to data scarcity. Recent data from Friedrich et al. (2020) show that 76% of global emissions in 2018 come from the energy sector. Zooming in on that sector shows that 91% of the emissions in the energy sector come from carbon dioxide which supports the definition of the corporate carbon footprint given in Wiedmann and Minx (2008).

To reach the Paris 2050 agreement, countries are obliged to reduce their corporate carbon footprint. The strategy of governments to meet the agreement resulted in a twentyfold increase in the number of climate change (related) laws since 2000 (Eskander and Fankhauser (2020); Nachmany et al. (2017)). The regulations include the obligation for large investors to disclose the Environment, Social and Governance (ESG) score of their investments. On the one hand, the ESG score is important to investors since Giese et al. (2019) found that investors include it in their valuation models

through their systematic- and idiosyncratic-risk profile. On the other hand, it is important to the management of companies too, since a high ESG score results in lower costs of capital and higher valuations (Ng and Rezaee (2015)).

Investors need to have standardized information available to fully incorporate ESG information in their valuation models. In order to make this information available, governments worldwide begin to pressure a growing number of companies to disclose standardized non-financial information. Some countries, such as China or South Africa, oblige companies that are listed on certain stock exchanges to disclose non-financial information (Ioannou and Serafeim (2017)). According to Ho (2020), the USA has regulations around the disclosure of non-financial information too, but the Securities and Exchange Commission (SEC) has not yet adopted any specific disclosure rules on ESG risks, nor do any of the current rules specifically mention environmental or social risk factors. The non-financial reporting directive (EU Directive 2014/95/EU) introduced in 2014 by the EU is planned to be updated in early 2023 with the objective to have a similar level of assurance for sustainability reporting as for financial reporting. Under this new directive, 4.5 times as many companies need to disclose non-financial information covering almost 50.000 companies in Europe instead of the 11.000 covered now. Next to the increase in companies, more information in a standardized reporting format needs to be disclosed, making it the most complete and strict directive at that time.

Companies need more insights into the sustainability performance of their company to meet the mandatory requirements. Already, companies are collecting data to report their impact on the environment. The availability of new standardized data can have a positive influence on the predictability of corporate carbon footprints in the future. Over the last twenty years, the increasing amount of data made it possible to create forecasting models. Previous studies made use of a naive sequence of extrapolation (Busch et al. (2020)) or performed a linear regression using Ordinary Least Squares (OLS) such as Goldhammer et al. (2016). However, while having more and more data, extra complexity is added to models when an increased number of variables and their interactions are incorporated. Non-linear dependencies between explanatory variables are possible, which makes OLS less sufficient. Nevertheless, some models are able to capture the increased complexity of the data. Machine learning models, such as ensemble methods, proved to be able to distinguish patterns within the data better than OLS when a lot of data is available (Re and Valentini (2012)). According to Ren et al. (2016), ensemble methods combine a series of machine learning models to improve the accuracy of the model substantially.

The ensemble method Gradient Boosting is introduced by Friedman (2001). It showed great

performance combining a high number of shallow trees, each modelling residuals of previous trees. An improved version, Extreme Gradient Boosting (XGBoost), is introduced by Chen and Guestrin (2016). In this ensemble method, multiple trees can be grown sequentially, decreasing the impact of individual trees on the final model. Different from gradient boosting, it introduces regularization parameters to reduce overfitting and uses random subset selection to increase inter-tree variance. Overfitting is a concept where a model may fit irregular (and unpredictable) noise of the training data making the model possibly less useful for out-of-sample forecasting. Nguyen et al. (2021) showed great prediction performance of this method forecasting the corporate carbon footprint.

Light Gradient Boosting Machine (LightGBM) is a recently developed gradient boosting framework that uses a tree-based learning algorithm. The biggest difference with other ensemble methods is that while other algorithms grow trees horizontally, LightGBM grows the trees vertically. This means that when growing the same leaf, the former can reduce more loss than the latter since the other ensemble methods need to keep the tree balanced by splitting all nodes on one level. LightGBM splits the node that reduces the most loss possibly making the tree unbalanced. Therefore, it is less computationally expensive than XGBoost (Ma et al. (2018)). Several dummy variables are introduced in this study to account for sectors, regions and industries. Including these dummy variables in the models result in sparse matrices. LightGBM handles sparse data matrices efficiently making it a possible best performing method in this study.

Next to ensemble methods, neural networks are able to deal with large complex data. Neural networks approximate the nonlinear function, but in a different semi-parametric manner than ensemble methods. The neural network specifies a global non-linear function using a flexible and layered structure, such that it can achieve highly accurate local approximations of the true relationship in the data. According to Gue et al. (2020), neural networks have shown better prediction performance compared to regression models in modelling complex behaviour of systems concerning the 17 Sustainable Development Goals.

Although the above methods showed great potential in other applications, they are never used to explain the relation between equity value and the corporate carbon footprint. Next to that, LightGBM is also never used to forecast the corporate carbon footprint of companies. The estimation of the corporate carbon footprint and the relation between equity value and the corporate carbon footprint will be the focus of this study.

This paper will contribute in multiple ways. First, an extension of machine learning techniques of Nguyen et al. (2021) will be explored to find better out-of-sample forecasting performance of the

corporate carbon footprint, where other papers mostly focus on in-sample fit. The use of machine learning in this field is new and this will be the second paper using these techniques to improve performance. Second, this paper will also engage in the widest possible universe of firms across sectors, regions and industries. Since the need for a general estimation method is growing, the estimation method must be able to be applied to all sectors, regions and industries. Third, the need of investors and regulators for more comprehensive estimates of scope 1 and 2 emissions is addressed since the emissions will be separately estimated.

Furthermore, this paper will be the first to use machine learning techniques to find relations between the corporate carbon footprint and equity value. If investors include negative valuations for emissions in their valuation models, they discount the equity value of companies that emit more emissions. To estimate this relation this study will zoom in into companies valued by a stock market index such as the S&P500. This gives the results economic relevance since the study provides empirical evidence concerning the extent to which investors incorporate often unassured, uncertain, non-financial information in their valuation models. It investigates the possibly non-linear relation between the corporate carbon footprint and equity value and estimates the price elasticity between the former two. Lastly, it enables companies to make well-informed strategic decisions because they can compare the impact of different strategies on their carbon footprint and equity value.

The data used in this research is based on the data set used in Nguyen et al. (2021) which contains data from 2005 to 2017. The data is gathered from the Thomson Reuters ESG universe that comprises over 8,500 global organizations across sectors, regions and industries. To include extra information, the data are complemented with static and continuous variables from the Eikon database. The models are trained on different sets of training data, after which out-of-sample data is used to test the actual prediction performance on unseen data.

All models used in this study have different hyperparameters that influence the performance of the respective model. These hyperparameters are optimized using a (sequential) gridsearch on the training data attaining the lowest mean absolute error. All out-of-sample predictions are compared on several performance metrics. This study makes use of the mean absolute error (MAE), the root mean squared error (RMSE) and the R-squared. To be able to determine if the prediction performance of models differ significantly the Diebold-Mariano test and the Model Confidence Set (MCS) are used. The results from OLS are highly interpretable in contrary to machine learning methods since coefficient estimates give a good insight into the relation between the dependent and the predictor variables. To find the most influential predictors in the machine learning models,

SHapley Additive exPlanation (SHAP) values are calculated. This method interprets the difference in impact on the dependent variable for a certain value of the predictor variable compared to a baseline value. Finally, the robustness of the methodology is determined. By choosing different numbers of subsets of the model the robustness can be determined. Furthermore, different out-of-sample performance metrics are compared.

This paper finds that LightGBM is the best estimation method for the corporate carbon footprint reducing the mean absolute error up to 13% compared to the benchmark OLS. Predictors that represent the scale of operations have the biggest positive impact on the size of the carbon footprint. The neural network estimates the equity value most accurate. Business-related predictors have the largest impact on equity value but a negative relation is found between equity value and scope 1 or scope 1+2 emissions. A sensitivity analysis shows that these relations are non-linear indicating varying discounts on equity value for different sizes of the carbon footprint. It also suggests a threshold where emissions above the threshold are not negatively discounted favouring heavy polluters.

The thesis is structured as follows. In chapter 2, the related literature is discussed. It gives an overview of previous research that has been done on the topic and the methodology. It exposes best practices and identifies gaps in current research. Next, the data is analyzed in chapter 3. It shows the process used to clean the data and how missing values are filled. Next to that, it displays several descriptive statistics of the data. Then the methodology is explained in Chapter 4. It analyzes the mathematical and statistical reasoning behind the methodology and explains how the models are build up. After defining the methodology, the results are discussed in chapter 5. All models are compared and the answer to the research questions is distilled from the information available. Lastly, chapter 6 gives a conclusion summarizing the most important findings. The discussion gives a critical reflection on the study and gives directions for further research.

2 Related Literature

The number of companies that disclose their emissions has increased tenfold in the period 2007-2018 (see Appendix A for the number of disclosers per year). However, still, not all companies disclose their emissions or only disclose certain scopes. This asks for a general estimation method to be able to estimate the corporate carbon footprints of companies in all sectors, regions and countries. In order to be able to design such an estimation method, relevant literature on current methods of measuring and estimation methods is discussed.

2.1 Current estimation methods

Several data providers such as Bloomberg, CDP, MSCI and Thomson Reuters provide data on carbon dioxide emissions. Their databases are based on publicly available yearly reports issued by companies themselves. See Appendix B for an overview of different methods that companies use to calculate the corporate carbon footprint. When data are not disclosed, estimation methods are used. Bloomberg and CDP do not estimate data contrary to Thomson Reuters and MSCI. According to Busch et al. (2020), their models are based on a naive sequence of extrapolation, emissions from comparable sectors or groups and historical emissions.

Other studies have developed estimation methods focusing on specific regions, sectors or industries. Goldhammer et al. (2016) use OLS to estimate scope 1 and 2 emissions from an external perspective. They focus on industries within Europe and use publicly available data as predictors. The best results are found combining sectors whilst adding sector-specific dummies to account for sector-specific emissions in one model. Griffin et al. (2017) focus on the American market, including S&P 500 companies in their analysis. They estimate a Gamma Generalized Linear Model, again using publicly available data. Where possible they first estimate direct and indirect emissions separately (scope 1 and 2) as a pooled, cross-sectional regression. If this is not possible they estimate the combined amount of emissions. A recent study uses a slightly different approach to forecast emissions. Nguyen et al. (2021) also use publicly available data but do not focus on a specific region. Using machine learning the study makes an estimation model which estimates scope 1, 2 and 3 emissions. It is the first study that covers the widest possible universe of firms across sectors, industries and regions. However, Nguyen et al. (2021) does not model the relation between equity value and the corporate carbon footprint.

Since regulators want to let the polluters pay for the size of their carbon footprint, it is important to quantify the costs of one ton of carbon dioxide emissions. Much prior research has been done studying the valuation effects of environmental disclosures. The findings have been twofold; Kolk et al. (2008) are doubtful of the relevance of carbon footprint disclosures in relation to equity value. On the other hand, more recent studies have shown negative relations between equity value and the corporate carbon footprint. Chapple et al. (2013) find a negative relation for Australian companies, Matsumura et al. (2014) and Griffin et al. (2017) for U.S. companies and Clarkson et al. (2015) for European companies.

These results agree on the negative relation between equity value and the carbon footprint. However, they all find a different negative valuation for the carbon footprint ranging from €11,-

per ton CO_2 for Australian companies to €180,- per ton CO_2 for U.S. companies[1]. The former studies focus on a specific region or sector and do not cover all companies. To be able to reach the Paris 2050 agreement all companies in the world need to pay for their pollution. In order to gain insights in the price investors put on the carbon footprint globally, a model needs to be made that is not limited by physical borders. Therefore a second model is designed to remove the limitations of previous research and to estimate the contemporaneous relation between equity value and different scopes of emissions.

2.2 Promising methods in similar studies

Including all original predictors in OLS is rarely effective. Removing redundant variables simplifies the model and can prevent overfitting. This study uses numerous predictor variables representing different characteristics of a company. However, some variables may be redundant since comparable variables are used to represent company statistics. Three methods are commonly used to prevent overfitting when using a least-squares method; Ridge, Lasso and Elastic Net. These shrinkage methods are linear methods making them highly interpretable. Especially for the model where the equity value is estimated the interpretability is of great value where the estimated coefficient shows the direct impact of the carbon footprint on equity value.

Neural networks were first proposed by McCulloch and Pitts (1943). However, only since Werbos (1982) introduced back-propagation it is a widely used method. The neural network can apply non-linear transformations to the data in order to model the parametric structure of the data (Bishop (1995)). Neural networks show improved performance compared to straightforward regression techniques in multiple finance and accounting studies according to Paliwal and Kumar (2009). Saleh et al. (2015) use a back-propagation artificial neural network to predict CO_2 emissions from boiler operations. Another study from Xu et al. (2019) finds superior predicting performance from a neural network compared to a nonlinear auto-regressive model predicting the CO_2 emission peak of China. Both studies show the potential predictive performance of neural networks in CO_2 related studies, however, they make use of non-public information where this study only uses publicly available information.

Multiple studies show the potential of ensemble methods. These methods combine the predictions from single models to one final prediction. Kadam and Vijayumar (2018) showed predictive performance of decision tree-based methods when predicting CO_2 emissions. Nguyen et al. (2021)

[1]A conversion rate of €0.62 per $1 Australian dollar is used. For U.S. companies, €0.85 per $1 is used.

showed that the random forest method showed a significant improvement in comparison with linear estimation methods studied earlier when forecasting the corporate carbon footprint. The study showed an even bigger improvement for the method Extreme Gradient Boosting (XGBoost). The method is introduced by Chen and Guestrin (2016) and they show that XGBoost outperforms other methods such as Random Forest and Neural Networks in different financial applications. The great performance follows from the ability to capture complex data dependencies and the fact that XGBoost is scalable and therefore capable of learning from large data sets. Light Gradient Boosting Machine (LightGBM) is a recently developed gradient boosting framework that uses a tree-based learning algorithm. LightGBM is less computationally expensive when the data is relatively sparse. This study includes several one-hot encoded dummy variables resulting in a partially sparse data matrix. Since Nguyen et al. (2021) found excelling prediction performance using XGBoost, LightGBM could reduce computational expenses and improve performance.

3 Data

This chapter describes the data used to answer the research questions. First, the companies that are included in this study are elaborated on. Second, the predictor variables are introduced and explained. Real-world data is used, so a proper data cleaning method, as well as a method to fill missing values, is introduced in the third part of this chapter. Finally, relevant data characteristics are shown.

3.1 Company selection

The data used in this research include companies worldwide covering a wide variety of industries and regions. The data set is based on the data used in the study of Nguyen et al. (2021) which use the ESG data set retrieved from Refinitiv[2]. This data set is provided by Refinitiv to help investors make an in-depth, responsible investment analysis based on multiple factors. Refinitiv measures a corporate's relative ESG performance, commitment and effectiveness based on publicly reported data. These factors include the return on investment, but also non-financially aspects such as environmental, social and governance performance. Their ESG universe comprises over 8,500 global companies, spanning major global and regional indices.

The entire data set before preprocessing contains 8,507 different companies with yearly observa-

[2]https://solutions.refinitiv.com/esg-data/

tions from 2007 to 2018. Note that Nguyen et al. (2021) uses observations from 2005 to 2017. The dataset includes companies that have gone bankrupt or closed their business in that period. It also includes companies that have been founded after 2007. These companies are included to prevent survivorship bias. Survivorship bias is the error when only companies are included that manage to stay active over a specific period of time which can lead to false conclusions.

3.2 Variable overview

3.2.1 Corporate carbon footprint model

Three different dependent variables are separately estimated by the various models. By splitting the carbon footprint into separate categories the model is able to predict direct and indirect emissions separately. This makes the model a potential estimation method for companies to create insights into their own footprint. It also enables regulators and investors to get a more detailed insight into the carbon footprint of companies. First, scope 1 emissions are estimated being direct emissions from owned or controlled company facilities. Scope 2 emissions are estimated representing indirect greenhouse gas emissions from purchased electricity, steam, heating and cooling for own use. Lastly, summed scope 1 and 2 emissions are estimated too. This study only focuses on scope 1, 2 and 1+2 emissions since there is little standardization or agreed degree of disclosure of scope 3 emissions. Also, the GHG Protocol developed by the World Resource Institute and World Business Council for Sustainable Development does not have a solution for double-counting issues in Scope 3 emissions (Institute and for Sustainable Development (2004)). Double-counting is the problem when certain emissions are included in two different scopes of emissions resulting in reporting the same emission twice. Because of this, investors and researchers are restricted to scope 1 and scope 2 emissions (Goldhammer et al. (2016); Griffin et al. (2017)).

This study uses the same predictor variables as assembled by the study from Nguyen et al. (2021) to predict the corporate carbon footprint. They carried out an extensive review of predictors used in earlier emission studies. They define five groups of variables namely, scale of operations, business model, technology, energy information and business environment.

A brief overview of the different variables is given in this section. See Nguyen et al. (2021) for a more extensive reasoning on the selection of the variables. This study includes six variables representing the scale of operations. First, the total annual revenue of a company is included since carbon emissions follow from revenue-generating processes. The earnings before interest, taxes depreciation and amortization (EBITDA) is included as well since it is useful to analyze companies

that are capital intensive. The number of employees is added next. It presumably influences scope 2 emissions, since a large number of employees need larger facilities resulting in higher use of electricity and heating. Next to that, having a high number of employees and low revenue may indicate a company where (raw) materials are processed hinting to higher scope 1 emissions. Next, the assets of a company are included via three variables. Total assets are included to account for the existence of large facilities. When the majority of the total assets are intangible assets this does not necessarily have to result in more emissions since the latter do not have direct emissions by definition. Physical assets, however, do result in more emissions. To account for this, the net plant, property and equipment (PPE) amount is included too. Lastly, a leverage variable is constructed by dividing the total debt by the total amount of assets to account for the financing strategy of a company.

Next, the business model is represented by the gross margin and industry type. The gross margin reflects the costs of goods sold in the value of the final product. A low gross margin possibly indicates more processing efforts hinting at higher emissions. Secondly, the Global Industry Classification Standard (GICS) industries are included representing 20 industries divided over 10 sectors (see Appendix D for an overview). To include the industries in the models, dummy variables are made using one-hot-encoding.

Companies around the world have different levels of technological advancement and can decide to prioritize investments in new technology. When a company exploits advanced equipment their emissions will possibly be lower compared to a company using old equipment since modern equipment is less polluting. A variable PPE age is made by dividing the gross PPE value by the depreciation indicating the age of the equipment used. High capital investments may result in more modern machinery with lower emissions as result. Lastly, the capital intensity is calculated by dividing gross PPE by revenue. Manufacturing companies have high capital intensity and higher emissions compared to service providing companies with low capital intensity.

To account for the different energy production methods over the world a fuel intensity variable is included measuring the carbon intensity of the national fuel mix. Pinpointing energy consumption to countries where factories and offices are located requires non-public company-specific information. Therefore, the carbon intensity of the national fuel combustion as reported by the International Energy Agency (IEA) is used for the country where the headquarter of a company is located in.

The last variable group represents the business environment of a company. Less developed countries tend to have a larger carbon footprint than richer, more developed countries. Since this study does not focus on a specific region where the business environment is constant a dummy variable is

included to distinguish three income groups as defined by the World Bank. The income groups are: 'high-income', 'upper-middle income' and 'lower-middle income'. Secondly, different countries have varying carbon-related laws (see Section 2). Companies located in a country with strict carbon-related laws will invest more in new technology to reduce their carbon footprint. Therefore, a dummy variable is included representing the presence of regulations as defined by the Worldbank. They define four categories: 'No CO_2 law', 'sub-national implemented', 'national implemented' and 'regional implemented'. Lastly, since companies try to reduce their emissions a yearly dummy is included to capture the negative trend.

3.2.2 Equity value model

There is only one dependent variable in the equity value model being the equity value of a company itself. The equity value constitutes the value of the company's outstanding shares multiplied by the market price. Data of different companies are combined in order to create a model that is able to estimate the equity value across regions, industries and sectors.

The predictors can be grouped into two categories namely, scale of operations and business environment where the scale of operations predictors are based on earlier work from Chapple et al. (2013), Matsumura et al. (2014) and Griffin et al. (2017). In this category, total assets are included since this is linked to equity value. Chapple et al. (2013) also includes the book value which represents the net value of a company's assets as reported on its balance sheet. The ratio between the equity value and book value is interesting. If the equity value is low compared to the book value this might indicate that the market values future growth negatively possibly suggesting non-compliance costs regarding ESG performance. Thirdly, the net income before taxes is included next to operating income to account for non-operating income, non-operating expenses, and other income. This type of non-operating income is assumed to have a little carbon footprint. So if this type of income has a relatively high impact on equity value it is accompanied by little emissions. Lastly, the total liabilities per year are included as liabilities have a negative impact on the equity value of a company.

The predictor group business environment contains a dummy to represent the presence of a CO_2 law. Lobbyists claim that implementing a CO_2 law can negatively impact the profitability of a company since they have to make extra costs contrary to companies that are not subject to this law. This implies that the equity value may be negatively impacted by the presence of a CO_2 law. To distinguish differences in industries the GICS industry dummies from the first model are included. Lastly, yearly dummies are included to account for a yearly trend in the market.

3.3 Data cleaning and missing values

The data set before pre-processing contains information on 8,507 companies. However, in this data set, some observations are of low quality. This subsection first describes the cleaning of the data set for both models. Secondly, it describes the methods used to fill missing values using different techniques following Nguyen et al. (2021).

3.3.1 Corporate Carbon Footprint model

First, this study filters data on the carbon footprint estimation method. The ESG universe from Refinitiv estimates carbon footprints if they are not disclosed by a company. Since this method adds error to the data, these observations are removed. Considering the vast increase in companies that disclose their carbon footprint, only the observation from the specific year for that specific company is deleted. Removing the entire company from the data set would entail an unnecessary loss of observations. See Appendix A to see the increasing number of companies that disclose their emissions over time illustrating the potential loss of information. After this step, the data contains 3,209 companies with 18,844 observations. Next, observations with missing scope 1 or 2 emissions are deleted since the supervised learning methods need a observation for the dependent variable leaving 18,235 observations. An extra variable Total Emissions is created by the aggregation of disclosed scope 1 and 2 emissions. Note that missing scope 3 values are treated differently since the focus of most research lays on scope 1 and 2 emissions. This focus follows from the fact that only scope 1 and 2 emissions are directly controlled by the company. Keeping that in mind, the data set is split into a set with no missing values for scope 3 emissions and one set where missing scope 3 emissions are ignored.

Then the predictor variables are analyzed. Observations that miss required financial data are deleted leaving 17,853 observations. This financial data is mandatory to disclose, so when a company did not disclose for instance reports of revenue, it indicates the unavailability of financial statements. To be able to distinguish the corporate carbon footprint between industries, industry classifications are necessary, so observations without are removed leaving 16,209 observations. Furthermore, observations with uncommon negative values for revenue, gross margin, asset age or capital intensities are deleted leaving 15,793 observations. Lastly, outliers have proven to negatively affect predictive performance. They can be present in the data because of different reasons such as calculation errors or mistakes in reporting. To cope with these outliers, observations that lie below the bottom 1st percentile and above the top 99th percentile are winsorized which is a common method in corporate

finance to cope with outliers (Mitton (2020)). Here, outliers are set to the 1st or 99th percentile changing 193 observations of the total sample. Finally, the missing values need to be handled in a proper manner following Nguyen et al. (2021). Missing values are first filled with the value of the same company in the next year. If this is not possible, the missing values are imputated using the mean of the companies in the same sector.

3.3.2 Equity value model

The data set used in the previous model is used as a basis. From this cleaned data set, observations for all companies on different scopes of emissions, information on CO_2 laws and GICS sector codes are imported. After including information on the equity value, total liabilities, total asset value, book value, operational income and the net income a data set with information on 2,646 companies including 14,683 observations is made. Next, observations that miss important financial data are again deleted resulting in a data set with 2,629 companies and 14,582 observations. 41 outliers are winsorized and the remaining missing values are filled using the method described in Section 3.3.1.

3.4 Data Characteristics

To be able to select the right method that fits both data sets best several data characteristics are explored which are displayed in Appendix E. This Appendix shows the number of observations that are in the different dummy categories as well as descriptive statistics on all variables. First, the skewness and kurtosis of the data is computed to check for asymmetric distributions. All continuous variables except the fuel intensity are highly skewed which is in line with findings in Goldhammer et al. (2016). This is to be expected since both samples are highly diversified in terms of types of companies, industries and regions. In order to reduce the skewness, the respective variables and dependent variables are log-transformed. Following Griffin et al. (2017) the original format of the ratio predictors leverage and gross margin are kept.

Different predictors are included in both data sets to represent company, industry and region characteristics. Since these variables can be highly correlated, some can be redundant possibly affecting the prediction performance negatively. In Appendix F it is shown that especially scale of operations predictors are highly correlated with coefficients > 0.5. Including these redundant variables in a standard linear regression can result in lower prediction performance. However, other methods included in this study have built-in predictor selection abilities alleviating the issue of redundant predictors to a certain degree as is shown in Section 4.

4 Methodology

This chapter describes the methodology used to answer the research questions. The framework used will be the same in the basis, but their respective hyperparameters will differ. First, the design of the train-test framework will be explained. After that, the methods will be introduced. To be able to compare the methods, several performance metrics are introduced. After the metrics are defined, the hyperparameter tuning process and robustness checks are described.

4.1 Design test framework

The linear models and the neural network in this study require scaled input data. Scaling transforms the data to have mean 0 and standard deviation 1. The scaling is performed using the following equation:

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}}, \tag{1}$$

where $\tilde{x}_{ij}$ represents the scaled observation i with $i = 1, ..., N$ and N equal to the total number of observations for predictor j. $\bar{x}_j$ is the mean of all observations x_{ij} for predictor j with $j = 1, ..., k$ and k equal to the number of predictors. Lastly, $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}$ represents the sample standard deviation.

To be able to measure the out-of-sample performance of a model a cross-validation approach is designed. The data set is split into training and test sets using group-K-fold-cross-validation. K-fold-cross-validation splits the data into K groups of equal size. The model is then trained on $K - 1$ training sets and tested on the K_{th} test set. This procedure is repeated K times where the test set shifts every iteration resulting in every fold being used as a training and a test set. Group-K-fold-cross-validation is an extension of the former where the same group will not appear in two different folds. Here, a group represents one company. The folds are roughly balanced in the sense that the number of unique groups is roughly the same in each fold. It is important to use this cross-validation method in this study because otherwise there would be information on the same company in the train and test set. When this is the case, the model is trained on for instance 5 years of data of one company also present in the test set resulting in false out-of-sample performance for the remaining 6 years of test data that are estimated. In practice, the value for K is often set to 5 or 10. Because of computational reasons, this study sets K equal to 5.

Different programming languages can be used to implement the methods. This research uses

Python to build the frameworks. Scikit-learn is used to implement group-K-fold-cross-validation. The linear models are build using the packages Linear Regression, Ridge, Lasso and ElasticNet from Scikit-learn. From the same library, the RandomForestRegressor is used to build the Random Forest. Lastly, this open-source package also provides the framework for the neural network using their MLPRegressor. XGBoost has an open-source ready to use package designed and provided by Chen and Guestrin (2016). LightGBM is also an open-source package provided by Ke et al. (2017) and Microsoft.

4.2 Linear Models

4.2.1 Ordinary Least Squares

OLS is one of the most popular statistical methods used in data analytics. OLS gives highly interpretable results since the relative influence of a predictor variable on the dependent variable can be read from the sign and size of the coefficient. Previous research on the corporate carbon footprint and its relation with equity value used OLS as estimation method. This makes OLS a good benchmark to compare the accuracy of other models too.

4.2.2 Shrinkage models

OLS is under standard assumptions unbiased. However, adding a little bit of bias can result in less variance and better results. Linear regression suffers from variance when many predictors are included in the model or when they are highly correlated which each other. Regularization allows reducing the variance at the cost of introducing some bias with the goal of reducing the total error of the model. Friedman (2001) showed that using different shrinkage techniques can improve prediction performance. Since Section 3.4 showed that some predictors are highly correlated shrinkage techniques could increase performance. Ridge regression as introduced by Hoerl and Kennard (1970) is a least-squares method with a L_2 constraint on the regression parameters, representing the Euclidean norm. Ridge is not able to select a predictor but shrinks the size of the predictor coefficients towards zero. In this way, the complexity of the model is continuously decreased, while keeping all predictors in the model

The Lasso, however, is able to shrink the coefficients of predictors to zero. The Lasso performs both variable selection and regularisation to improve both prediction accuracy and interpretability of a model. This is achieved by adding a L_1 penalty term to the OLS framework shrinking the

coefficient of one single predictor to zero. This is beneficial when only a few predictors have predicting performance which can be the case in this study since potential redundant predictors are included. Lasso will be able to distinguish the predictors that have the most impact on the dependent variable.

Since both Lasso and ridge have their limitations another method is introduced. The elastic net is a convex combination of the penalty parameter L_1 and L_2 and is therefore seen as a combination of Lasso and ridge regression combining their unique advantages.

All shrinkage methods aim to minimise the following residual sum of squares (RSS) with an additional penalty term:

$$\text{RSS} = \underset{\beta \in \mathbb{R}^k}{\arg\min} \left\{ ||Y - X\beta||_2^2 + \alpha \ \lambda_1 \sum_{j=1}^{k} ||\beta_j||_1 + \alpha \ (1 - \lambda_1) \sum_{j=1}^{k} ||\beta_j||_2^2 \right\}, \tag{2}$$

where Y $[N \times 1]$ represents a vector with scope 1, 2 or combined scope 1 and 2 emissions in the model where the corporate carbon footprint is estimated. The vector has length N with N equal to the number of yearly observations for the companies in the data set. In the other model, Y is a $N \times 1$ vector representing the yearly equity value of N companies over the period 2007-2018. Next, X $[N \times k]$ represents the matrix with k predictor variables. The estimated regression coefficients β $[k \times 1]$ are calculated as $\hat{\beta} = (X'X)^{-1}X'Y$. The L1 penalty term is represented by $\lambda_1 \sum_{j=1}^{k} ||\beta_j||_1$ and the L2 penalty term by $(1 - \lambda_1) \sum_{j=1}^{k} ||\beta_j||_2^2$.

A hyperparameter that can be optimized is λ_1 influencing the Lasso/ridge shrinkage ratio. When λ_1 is equal to 0, equation 2 represents ridge shrinkage, when λ_1 is equal to 1 it represents Lasso shrinkage. Hence, a gridsearch is performed for values for λ_1 from 0.01 to 1. A higher penalty term α results in more shrinkage, however, when the penalty term is too high, all coefficients are shrunken towards zero. To determine optimal α a gridsearch is performed for multiple values between 0.0001 and 100. See Appendix G for the process that uses cross-validation and a gridsearch to determine the combination of hyperparameters that result in the highest prediction performance.

4.3 Ensemble methods

The idea behind ensemble methods is to combine several base models in order to create one optimal predictive model. This can be done using several techniques which will be discussed in the following subsections. By combining multiple models and averaging their prediction, variance is reduced at the cost of the introduction of a little bias.

The ensemble methods used in this study are tree-based methods that try to divide the pre-

dictor space into a number of simple regions based on feature values. Tree-based models are the most popular and relevant methods used today. They can be used for both classification and regression problems. Corporate carbon footprints and equity values are continuous values, so regression methods are most suited here. Regression methods find regions such that the objective function is minimized. The maximum number of regions is equal to the number of observations in the data. If all observations have their own region, the model overfits the data leading to worse out-of-sample performance. To overcome this problem, a stopping rule is implemented. Different stopping rules can be used such as a minimum number of observations per region or a maximum number of regions.

4.3.1 Random Forest

Random forest is a very popular learning method because of its properties. It is simple to understand and easy to implement. It can handle non-linearity well which may be present in both data sets. Finally, it is able to train in parallel and it can handle large amounts of data fast.

The goal of a random forest is to de-correlate the trees using bagging, without increasing the variance too much. The random forest grows a 'forest' of independently build shallow decision trees. It is called random since all trees are grown using bootstrapping. Bootstrapping makes subsets using random sampling with replacement. Hereafter, separate shallow trees are grown on the different subsets and the predictions are averaged to reduce variance. By averaging trees, the noise of the estimation is reduced greatly. However, since the final result is a combination of many trees the interpretability goes down as well. The prediction of the random forest is defined as follows:

$$\hat{y}^{RF}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{f}_b(x), \tag{3}$$

where B is the number of grown trees in the forest and $\hat{f}_b(x)$ is the prediction obtained with the b^{th} tree. To be able to determine which splits need to be made in the tree in order to make the prediction, a loss function is implemented. Every iteration, this function is optimized to find the split that maximizes the decrease in the loss function. This study uses the standard Residual Sum of Squares (RSS) as optimization function given by:

$$\text{RSS} = \arg\min_{\beta \in \mathbb{R}^k} \left\{ ||Y - X\beta||_2^2 \right\}, \tag{4}$$

To reach the potential of a random forest certain hyperparameters are tuned. Since a large data set with numerous predictors are used in this study, overfitting is a possibility. This study

first defines the maximum depth of the trees to prevent trees from growing too deep. When a tree grows too deep it can overfit the data. Next, the minimum number of samples per split is set as a hyperparameter. This prevents overfitting too since it sets the minimum number of samples required to split an internal node decreasing the number of splits if the minimum number increases. The minimum number of samples required to be at a leaf node determines if a split is considered or not. If the number of samples in the left and right node exceeds the minimum, the split is considered. Again, this prevents overfitting limiting the number of possible splits. The hyperparameters are tuned using sequential gridsearches. Here, the grid of Nguyen et al. (2021) is expanded with higher values for the hyperparameters to optimize performance because their optimal results were found for hyperparameter values at the end of their grid. The results of this approach can be found in Appendix G.

4.3.2 Extreme Gradient Boosting

Where a random forest grows numerous shallow trees simultaneously, a gradient boosting method relies on the principle of repeatedly improving a shallow model learning from the error made by previously trained trees. It adds the modelled residuals to the original model to increase the weight of the wrongly classified samples in the updated model. XGBoost can grow multiple trees sequentially making it an efficient and scalable variant of the gradient boosting algorithm (Chen and Guestrin (2016)). XGBoost is tested in many corporate finance benchmark studies and it has won many machine learning prediction competitions. Nguyen et al. (2021) found that XGBoost was the best performing prediction method estimating the corporate carbon footprint.

XGBoost is designed to be highly efficient. It runs very fast while introducing regularization parameters to reduce overfitting. Next to regularization parameters, it introduces column subsampling where random subsets of predictor variables are selected in each sequence of gradient boosting. Since this study makes use of highly correlated predictor variables randomly selecting predictors can improve model performance. Because of the random selection, it highly encourages variance between different trees allowing the model to converge faster. This study uses one-hot-encoding of categorical predictors so large sparse matrices are used as input. XGBoost is able to handle data that is sparse, has missing values or zeros in an efficient matter.

According to Chen and Guestrin (2016), XGBoost uses K additive functions to predict the output:

$$\hat{y}_i^{XGB} = \phi(\mathbf{x}_i) = \sum_{k=1}^{K} f_k(\mathbf{x}_i), \quad \text{with } f_k \in \mathcal{F}, \tag{5}$$

where $\mathcal{F} = \left\{f(\mathbf{x}) = w_{q(\mathbf{x})}\right\} \left(q : \mathbb{R}^m \to T, w \in \mathbb{R}^T\right)$ is the space of regression trees, T is the number of leaves in the tree and q represents the structure of each tree. f_k refers to an independent tree q with leaf weights w. Next, the objective function to be minimized is:

$$\mathcal{L}(\phi) = \sum_{i=1}^{N} l\left(\hat{y}_i^{XGB}, y_i\right) + \sum_{k=1}^{N} \Omega(f_k), \qquad \text{with } \Omega(f) = \gamma T + \tfrac{1}{2}\lambda\|w\|_2^2, \tag{6}$$

where l is a differentiable convex loss function that measures the bias of the predictions $\hat{y}_i^{XGB}$ for $i = 1, ..., N$ and Ω reduces the complexity of the model. The regularization term smoothens the final weights w in order to prevent overfitting.

Since equation 6 has functions as parameters it cannot be optimized in Euclidean space using traditional optimization methods. To cope with this problem, XGBoost trains the model in an additive manner. We let $\hat{y}_i^{(t)}$ be the prediction $\hat{y}_i^{XGB}$ of the i$^{\text{th}}$ instance at the t$^{\text{th}}$ iteration. Next, we greedily add the f_t that is the best improvement to the model according to equation 6. A second-order Taylor approximation is used to optimize the greedy function:

$$\mathcal{L}^{(t)} \simeq \sum_{i=1}^{n} \left[l\left(y_i, \hat{y}^{(t-1)}\right) + g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i) \right] + \Omega(f_t), \tag{7}$$

where $g_i = \partial_{\hat{y}(t-1)} l\left(y_i, \hat{y}^{(t-1)}\right)$ and $h_i = \partial^2_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}^{(t-1)}\right)$ are first and second order gradient statistics of the loss function. Consult Chen and Guestrin (2016) for a fixed structure of $q(x)$, the optimal weight w_j^* of leaf j and the corresponding optimal value of $\tilde{\mathcal{L}}^{(t)}(q)$. These first and second order gradients are the difference with the original gradient boosting algorithm which only incorporates a first order gradient. XGBoost also introduces a regularization term as shown in equation 6 to prevent overfitting.

Hyperparameters are tuned using a sequential gridsearch. First, the maximum depth of the trees grown is tuned to prevent overfitting. A relatively small grid is chosen here since Nguyen et al. (2021) showed consistent results in their hyperparameter optimization. Secondly, the minimum weight of the child nodes is optimized which determines the minimum amount of samples that need to be in the subspace at a child node. Next, the sub-sample ratio of training instances is determined which sets the ratio of training data that is randomly sampled prior to growing trees to prevent overfitting. The sub-sample ratio of the columns when constructing each tree is also tuned. Since some predictors are highly correlated the sub-sample grid spans 0.5 to 1 where a value of 0.5 means

that only 50% of the predictors is sampled. Finally, the learning rate is optimized. A higher learning rate means a larger step size in each update to prevent overfitting. After each boosting step, the weights of new features are calculated. The learning rate shrinks these feature weights to make the boosting process more conservative. Since the learning rate influences computational time but also has a great impact on the convergence of the model a large grid is defined for the learning rate. See Appendix G for the precise grids and results of the sequential gridsearch.

4.3.3 Light Gradient Boosting Machine

Light Gradient Boosting Machine (LightGBM) is another recently developed iteration of the gradient boosting framework. Where XGBoost includes a second derivative in the optimization function, LightGBM uses a different alteration. Instead of growing trees horizontally, LightGBM grows trees vertically. A tree is thus grown leaf-wise instead of level-wise as shown in Figure 1. Here, the above growing process shows the level-wise grow method used by XGBoost. The level-wise strategy keeps the tree balanced by splitting all nodes on one level. However, the leaf-wise reduces more loss by splitting the leaf that has the most loss possibly making the tree unbalanced as is shown in the leaf-wise growth process in Figure 1. An advantage when growing the same leaf is that the latter can reduce more loss than the former making it less computationally expensive as XGBoost (Ma et al. (2018)). A faster algorithm is beneficial for investors comparing numerous companies since this process can then be more time-efficient.

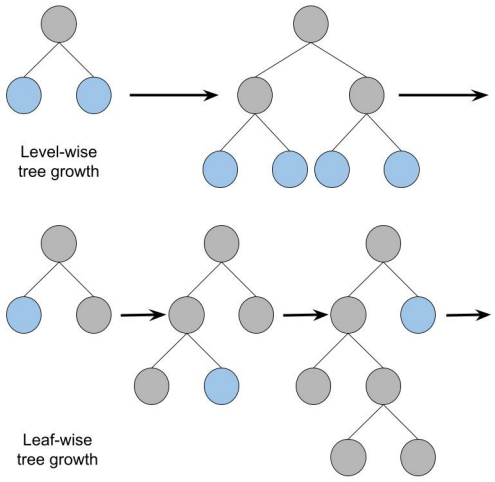


Figure 1: Leaf versus level-wise tree growth

LightGBM is a type of gradient boosting decision tree (GBDT) that is build to be used in

mass data. Both data sets in this study are quite large so both models benefit from this property. Conventional implementations of GBDT need to scan all data instances for every variable to estimate the information gain of all possible split points while LightGBM chooses a different approach. LightGBM aims to minimize a specific loss function $\mathcal{L}(y, f(x))$ as follows:

$$\widehat{f} = \arg\min_{f \in \mathcal{F}} E_{y,x} \left[\mathcal{L}(y, f(x)) \right], \tag{8}$$

where $\mathcal{L}(y, f(x))$ is equal to the loss function shown in Equation 6 as used by XGBoost and $\mathcal{F}$ is the space of regression trees as in Equation 5.

For GBDT, following Ke et al. (2017), the split at a node is usually based on the information gain measured by the variance after splitting. Another way of sampling is to use the information gain to split the data. However, classic methods based on weights can not be applied since there is no sample weight in GBDT. Ke et al. (2017) propose to use gradients to get insights into the information gain of a sample. They first rank the training instances according to the absolute values of their gradients in descending order. Second, the highest $a \times 100\%$ instances are gathered in a subset A. From the remaining set A^c, a random sample B is chosen randomly with size $b \times |A^c|$. A small gradient implies a small training error, while a large gradient implies a large training error resulting in a larger information gain. Since the information gain of the small gradients is minor they can be eliminated. As a result, the distribution of the data is changed negatively affecting the accuracy of the model. To overcome this, gradient-one-side-sampling is used. Here, the instances are split according to the estimated variance gain $\tilde{V}_k(d)$ over the subset $A \cup B$:

$$\tilde{V}_k(d) = \frac{1}{n} \left(\frac{\left(\sum_{x_i \in A_l} g_i + \frac{1-a}{b} \sum_{x_i \in B_l} g_i \right)^2}{n_l^k(d)} + \frac{\left(\sum_{x_i \in A_r} g_i + \frac{1-a}{b} \sum_{x_i \in B_r} g_i \right)^2}{n_r^k(d)} \right), \tag{9}$$

where A_l and A_r represent the data in subset A that is split according to predictor k at point d into left and right child nodes. B_l and B_r represent a similar split for subset B. g_i represents the negative gradients of the loss functions with respect to the output of the model for $i = 1, ..., N$ with N equal to the yearly-company observations. n_l^k and n_r^k represent the number of observations x_{ik} that are respectively in the left and right child node. The coefficient $\frac{1-a}{b}$ is used to normalize the sum of the gradients over B back to the size of A^c.

By using the estimated variance over a smaller subset, the algorithm is computationally less expensive while the loss of training accuracy is kept to a minimum (Ke et al. (2017)). Gradient-one-side-sampling makes the model focus on the data with large errors whilst keeping the data

distribution almost the same as the original data distribution. In this way, investors and other users can use the algorithm efficiently whilst attaining high accuracy making it a practically deployable method.

The categorical variables in the data need to be one-hot-encoded to be used in LightGBM. The datasets in this study have numerous categories so one-hot-encoding results in large sparse matrices with numerous sparse features. These sparse features are often mutually exclusive meaning that they do not have the same non-zero observations. The LightGBM algorithm bundles features that (almost) never take nonzero values simultaneously into a single feature using a greedy algorithm. This process is called Exclusive Feature Bundle (EFB) with the goal of reducing the dimension. Ke et al. (2017) showed that the same feature histograms from the EFB can be built as those from individual features. In this way, the complexity of histogram builds changes from a $N \times k$ to a $N \times l$ problem with l equal to the number of bundles. Since $l << k$, the training of GBDT can significantly speed up without losing much accuracy again advocating the practical use of LightGBM.

LightGBM has multiple parameters that can be tuned. First, the maximum number of tree leaves for base learners is tuned using a grid between 100 and 1500 leaves. This is the main parameter to control the complexity of the tree model. Using a leaf-wise growth strategy results in potentially deeper trees compared to a level-wise growth strategy while using the same number of leaves (Alshari et al. (2021)). This characteristic has the effect that the same value for the maximum depth hyperparameter from XGBoost results in trees with different levels of complexity for LightGBM. Analyzing both parameters may indicate differences in the complexity of the trees grown by XGBoost and LightGBM. Next, the minimum amount of samples needed in one leaf is tuned where a higher value decreases the possibility of overfitting since fewer sub-regions can be made. The sub-sample ratio of the training samples that will be used to train each tree and the sub-sample ratio of features that will be used when constructing each tree is tuned using a grid from 50% of the features to 100%. Since some predictors are highly correlated taking a subset may improve performance. Finally, the learning rate is determined. It determines the impact of each tree on the final outcome. LightGBM updates its initial estimate using the output of the tree. A larger learning rate equals a larger magnitude of change. On the one hand, a too big learning rate can result in divergence. On the other hand, a too-small value can lead to overfitting and a computationally expensive model. Here, different values are evaluated in a grid ranging from 0.001 to 0.3. Consult Appendix G to see the results from the sequential gridsearch.

4.4 Neural Networks

Over the last few decades, much more data has become available. Due to technological inventions, much more computational power is available now too. This plays an important role in the recent revival of machine learning since more complex, more flexible network architectures such as deep learning are usable now. An example of a deep learning method is a neural network. A neural network is inspired by the functionality of our brain where billions of interconnected neurons process information in parallel. Information is transferred via synapses from axons to dendrites. A single neuron may be connected to many other neurons and the total number of neurons and connections in a network may be extensive. This layered structure is mimicked by a neural network.

The neural network shown in Figure 2 consists of a network of nodes that is grouped in layers to mimic the above structure of the brain. The nodes in the hidden layer receive their input values from the nodes in the input layer. In these hidden layers, mathematical functions are applied to the received values. Hereafter, the results are transferred from the hidden layers to the output layer. In the output layer, each node represents a target variable that the model attempts to fit. Similar to the hidden layers, the output layer also applies a mathematical function to the input values. The result from these calculations is the output of the neural network for each dependent variable. After each iteration, the mathematical functions are changed based on the calculated error of the fitted values in an attempt to minimize the error (Bishop (1995)).

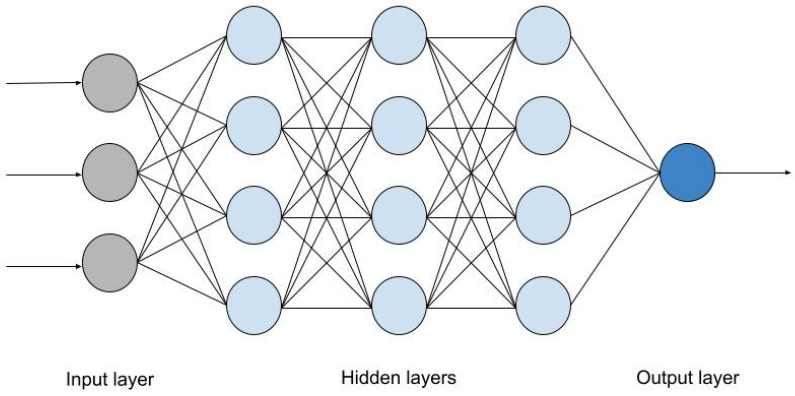


Figure 2: Neural network example

The input layer consists of k nodes, where k is equal to the number of predictors. The nodes in the input layer are connected to all nodes in the first hidden layer. In this hidden layer, mathematical functions are applied to the input values α_p and these functions are called activation functions. These activation functions may include logistic sigmoid, hyperbolic tangent, rectified linear unit, a linear activation function and many iterations of the before mentioned functions. These activation functions can model several data characteristics such as non-linearity.

Since non-linearity could be present in both data sets, the rectified linear unit (ReLu) is included as a possible activation function for the hidden layer. This activation function is commonly most successful and widely used according to Ramachandran et al. (2017):

$$f(Z) = \max(0, Z), \qquad \text{where } Z = w_{0k} + \sum_{j=1}^{k}(X_j w_{j,m}). \tag{10}$$

Here, $w_{0,k}$ is a bias coefficient, X is a vector with N observations for predictor k and $w_{j,m}$ is a set of weights for m nodes for j yearly-company observations. The weights are updated every iteration in order to create a model that fits the training data most accurately without overfitting it. A neural network is able to 'deactivate' nodes using the bias decreasing computational cost. It uses a threshold to determine whether the node is used or not and every node has its own bias coefficient. The output sent to the next layer is by definition non-negative. This next layer can be another hidden layer where similar activation functions are used again or the output layer where another function determines the output value. Unfortunately, adding more hidden layers decreases the interpretability of the model.

In the output layer of a classification problem, every possible category has its own node. However, since this study covers a regression problem the output layer only consists of one node representing a dependent variable. Hence, for every dependent variable, a neural network is estimated. The result from the last hidden layer is the input of the output layer. Here, a similar mathematical transformation is applied as in a hidden layer, however using a different activation function. For the output function, the linear activation function is used since it is a regression problem giving the following estimation for the dependent variable:

$$\hat{Y}^{NN} = \gamma_0 + \sum^{k} f(Z)\gamma_k, \tag{11}$$

where γ_0 is the bias, γ_k represent the weights and the hidden layers Z are combined using the linear activation function f to compute an approximation of the dependent variable.

To be able to find the optimal weights and biases an optimization function is defined. This study uses the squared error loss function which is computed every iteration at the output node. Afterwards, back-propagation is applied to update the weights and biases iteratively. In traditional gradient descent methods, each update of the parameter estimates is based on the gradient of all N observations in the training sample which is too computationally intensive to use in this study. Therefore, stochastic gradient descent is used where the update is based on the gradients of a random subset of the observations. Each iteration, the gradient is used to update the weights and biases in order to minimize the optimization function. The training of the model stops when there is no significant improvement anymore.

Since every layer can have a different activation function, possible non-linear relations in the data can be modelled. Combining multiple layers makes a neural network very flexible but decreases the interpretability of the model. This makes the neural network less practical to estimate the equity value model in this study. To be able to open the 'black-box' of the neural network other techniques are needed that may be unsuited for policymakers, companies and investors. The corporate carbon footprint estimation model focuses more on prediction performance. Therefore, the lack of interpretability is of less importance there. Increasing the number of predictors and number of layers can also increase the accuracy of the fitted model. However, adding too many can lead to overfitting. The flexible nature of the neural network makes that the model often is overfitted. To cope with this potential problem several methods are implemented.

Learning rate shrinkage is implemented to prevent overfitting. A high value for the learning rate equals a large step size between iterations. Initially, a large step size is desirable to speed up the optimization process. However, a too large step size can result in divergence instead of convergence to a local minimum. A too-small step size equals the need for many iterations to converge to a minimum making it computational too expensive. Adam is a method that combines the former two first introduced by Kingma and Ba (2014). It uses a large learning rate in the beginning and a smaller learning rate later in the process. Since all methods have pros and cons a constant, inverse scaling (adam) and adaptive learning rate are implemented as hyperparameters using a gridsearch. The adaptive learning rate starts constant but becomes smaller when two consecutive epochs fail to decrease training loss by at least a threshold of 0.0001. Here, an iteration equals one epoch which means that each data point will be used once per epoch.

To optimize the neural network other hyperparameters are tuned too. First, the number of nodes in the hidden layers are tuned. The hyperparameters cover neural networks with one and two hidden

layers each having up to 100 nodes. Next, the value for alpha is tuned. Alpha is an L2 regularization term that prevents overfitting by putting constraints on the size of the weights. Higher values for alpha may fix high variance by encouraging smaller weights resulting in less overfitting. Decreasing the value of alpha may fix high bias by encouraging larger weights resulting in a more flexible model. A uniform grid between 0 and 1 is tested in the optimization. Finally, the maximum number of iterations is altered. The model iterates until convergence or the maximum number of iterations are reached.

4.5 Evaluation criteria

To be able to tell which method is able to forecast the corporate carbon footprint and equity value the best, several evaluation criteria are defined. First, the performance of a specific method is evaluated. Later, the methods are compared against each other to see if there is a significant difference between models.

4.5.1 Root Mean Squared Error

First, this study implements the Root Mean Squared Error (RMSE). It is a quadratic scoring measure which computes the average magnitude of the error. This metric represents the accuracy of the forecasts and is computed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}, \tag{12}$$

where $\hat{y}_i$ denotes the forecasted scope 1, 2, combined emissions or equity value. Since the errors are squared relatively larger weights are given to errors with a large magnitude. Hence the RMSE favours a model which does not have large errors. This can be an interesting metric since it could be possible that a policymaker wants to trade a little bit of accuracy for a model that has no large errors.

4.5.2 Mean Absolute Error

The Mean Absolute Error (MAE) is to some extend comparable to the RMSE. The MAE measures the average magnitude of the errors without considering their direction. The difference with RMSE is that the MAE is a linear scoring method meaning that all errors are equally weighted in the summation. The MAE is calculated as follows:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|, \tag{13}$$

where $\hat{y}_i$ denotes the forecasted scope 1, 2, combined emissions or equity value. The greater the difference between the RMSE and MAE, the greater the variance in the individual errors in the sample. Policymakers could use this difference to see if specific sectors have greater variance in their errors. This is an indication that these specific sectors are harder to estimate because of for instance lower quality of data.

4.5.3 R-squared

Finally, the out-of-sample (OOS) R-squared is calculated. It is a metric to test whether the method has OOS predictability. It measures the fraction of the variation that is explained by the method of interest and the formula is given by:

$$\text{R}^2_{OOS} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2}, \tag{14}$$

where $\hat{y}_i$ denotes the forecasted scope 1, 2, combined emissions or equity value.

4.5.4 Model Confidence Set

In order to differentiate statistically between the forecasts, this study uses the Model Confidence Set (MCS) as introduced by Hansen et al. (2011). The MCS is typically used to compare a substantial number of models. However, there are previous studies such as Shang and Haberman (2018) that successfully implemented MCS using smaller sets of models.

Hansen et al. (2011) consider a set, $\mathcal{M}^0$, that contains a finite number of models for $i = 1, ..., m_0$ with m_0 equal to the total number of models compared. The models are evaluated in terms of a loss function and the loss for model i and observation N is given by $L_{i,t}$ for $t = 1, ..., N$. We denote $d_{ij,t} \equiv L_{i,t} - L_{j,t}$ for all $i, j \in \mathcal{M}^0$ and assume that $\mu_{ij} \equiv \text{E}(d_{i,t,t})$ is finite and does not depend on t for all $i, j \in \mathcal{M}^0$. Next, alternatives are ranked in terms of expected loss, such that alternative i is preferred over alternative j if $\mu_{ij} < 0$.

The objective of the MCS procedure is to determine $\mathcal{M}^*$ which is the set of superior objects defined by:

$$\mathcal{M}^* \equiv \left\{ i \in \mathcal{M}^0 : \mu_{ij} \leq 0 \text{ for all } j \in \mathcal{M}^0 \right\}. \tag{15}$$

This is done through a sequence of significance tests, where objects that are less significant than other elements of $\mathcal{M}^0$ are eliminated. The MCS is defined as the subset of $\mathcal{M}^0$ that contains all of $\mathcal{M}^*$ with a given coverage probability. The hypotheses that are tested are:

$$H_{0,\mathcal{M}} : \mu_{ij} = 0 \quad \text{for all } i,j \in \mathcal{M}, \qquad \text{where } \mathcal{M} \in \mathcal{M}^0. \tag{16}$$

The alternative hypothesis, $\mu_{ij} \neq 0$ for some $i,j \in \mathcal{M}$ is denoted by $H_{A,\mathcal{M}}$. The procedure is based on an equivalence test, $\delta_{\mathcal{M}}$, and an elimination rule, $e_{\mathcal{M}}$. The equivalence test is used to test the null hypothesis $H_{0,\mathcal{M}}$ for any $\mathcal{M} \in \mathcal{M}^0$. The object of $\mathcal{M}$ that is removed from the former is identified by $e_{\mathcal{M}}$ in the event that the null hypothesis is rejected. We let $\delta_{\mathcal{M}} = 0$ and $\delta_{\mathcal{M}} = 1$ correspond to the cases where $H_{0,\mathcal{M}}$ are accepted and rejected, respectively. The MCS algorithm is given by:

Algorithm 1: Model Confidence Set Algorithm

(a) Initially set $\mathcal{M} = \mathcal{M}^0$.

(b) Test $H_{0,\mathcal{M}}$ using $\delta_{\mathcal{M}}$ at level α

(c) **if** $H_{0,\mathcal{M}}$ *is accepted* **then**
| define $\widehat{\mathcal{M}}^*_{1-\alpha} = \mathcal{M}$

else
| use $e_{\mathcal{M}}$ to eliminate an object from $\mathcal{M}$ and repeat the procedure from step b

end if

Result: $\widehat{\mathcal{M}}^*_{1-\alpha}$

$\widehat{\mathcal{M}}^*_{1-\alpha}$ contains the set of models that are not eliminated and is referred to as the MCS. This means in practice, that if there is one model in the MCS it means that it is the best performing model given a level of confidence α. However, when multiple models are present it indicates that no significant difference in performance is present within these models in the MCS. In this scenario, an investor or policymaker could argue to use a model with insignificantly lower accuracy in favour of for instance interpretability.

4.6 Predictor Importance

The linear methods used are highly interpretable since the coefficients represent the relative importance of the predictors to the dependent variable. A single decision tree is highly interpretable too when the maximum depth is not too large. However, a known downside of ensemble methods is the drop in interpretability since multiple trees are combined. The neural network is known as

a black-box model where the relative importance of predictors is typically unknown. This lack of interpretability is especially a disadvantage for the model where the relation between the corporate carbon footprint and equity value is estimated. To be able to quantify this relation, insights into the methods need to be gained. In this section, a method that shows the importance of predictor variables is explained for the ensemble methods and the neural network.

This study uses multiple types of methods so a generally applicable approach to interpret model predictions is beneficial. Lundberg and Lee (2017) introduced an algorithm to reverse-engineer the output of any predictive algorithm called SHapley Additive exPlanations (SHAP). It is based on a form of game theory introduced by Shapley (2016) where Shapley values quantify the contribution of each player to a game. Here, the game is reproduced by the model's outcome and the players are represented by the predictor variables. Shapley quantifies the contribution of each player to the game, SHAP quantifies the contribution of each feature to the prediction made by the model.

SHAP values have several benefits to explain the output of any machine learning method. Firstly, SHAP values show how much each predictor contributes to the target variable, either positively or negatively. The feature importance used in Nguyen et al. (2021) is only able to tell the importance rather than the sign of the contribution. Secondly, each observation gets its own SHAP values showing how predictors contribute on a local level. This local interpretability makes it possible to pinpoint and contrast the contributions of the predictors.

To be able to compute the SHAP values, this study denotes some notation. The set containing all predictor variables is represented by F. The subset of predictor variables of a given trained model is represented by S such that $S \subseteq F$. Since the effect of retaining a predictor depends on other predictors in the model, the differences are computed for all possible subsets $S \subseteq F \backslash \{i\}$. The SHAP values can be computed according to Lundberg and Lee (2017) using:

$$\phi_i = \sum_{S \subseteq F \backslash \{i\}} \frac{|S|!\,(|F|-|S|-1)!}{|F|!} \left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S\left(x_S\right) \right], \tag{17}$$

where a model $f_{S\cup\{i\}}$ is trained with the predictors present in subset S and another model f_S is trained without these predictors. Next, the predictions from the two models are compared on the current input using the second part of Equation 17 given by $f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S\left(x_S\right)$, where x_S represents the yearly-company observations for the predictors present in subset S. As can be seen in Equation 17, the SHAP values ϕ_i are the weighted average of all marginal contributions of a specific feature. Analyzing these SHAP values give valuable insights that are discussed in Section 5.

4.7 Sensitivity analysis

The Partial Derivative (PaD) method is implemented to explore a potential non-linear relation between equity value and the carbon footprint (Lu et al. (2001)). The sensitivity can be expressed as the first-order partial derivative between the output variable and the input predictor. The gradients are used to make a graph of the output variations for small changes of each input variable to see if the relation is (non-)linear. A non-linear relation implies different behaviour for different sizes of carbon footprints. For instance, an above-average large carbon footprint could result in an extra valuation discount.

The neural network consists of multiple interconnected layers where linear transformations are combined (see Figure 2). The derivation of the first-order derivative is based on the derivations in Nourani and Fard (2012). The input data X_i is transferred in the input layer without any processing:

$$N_i = X_i, \tag{18}$$

The input is then passed to the neurons in the hidden layer where the activation function is applied to a weighted combination of the input data:

$$\begin{aligned} S_h &= N_p W_{hp} + \sum_{i \neq p} N_i W_{hi}, \\ N_h &= \phi_h \left(S_h \right), \end{aligned} \tag{19}$$

where p in Equation 19 represents the input predictor which impact is analyzed. The output from neuron h in the first hidden layer is in a 1-layer model transferred to the output neuron:

$$\begin{aligned} S_o &= N_h W_{oh} + \sum_{j \neq h} N_j W_{oj}, \\ \hat{Y} &= \phi_o \left(S_o \right), \end{aligned} \tag{20}$$

where j represents the neurons in the hidden layer other than h. The first-order partial derivatives over the predictor X_p is:

$$\frac{\partial \hat{Y}}{\partial X_p} = \frac{\partial \hat{Y}}{\partial N_p} = \frac{\partial \hat{Y}}{\partial N_h} \frac{\partial N_h}{\partial N_p} = \left(\frac{\partial \hat{Y}}{\partial S_o} \frac{\partial S_o}{\partial N_h} \right) \left(\frac{\partial N_h}{\partial S_h} \frac{\partial S_h}{\partial N_p} \right), \tag{21}$$

where

$$\frac{\partial \hat{Y}}{\partial S_o} = \phi_o' \left(S_o \right) \quad \text{and} \quad \frac{\partial N_h}{\partial S_h} = \phi_h' \left(S_h \right) \tag{22}$$

giving:

$$\frac{\partial \hat{Y}}{\partial X_p} = \phi_o' \left(S_o\right) W_{oh} \phi_h' \left(S_h\right) W_{hp}. \tag{23}$$

Here, $\phi'(.)$ represents the first-order derivative of the activation function used in the neural network. See Section 4.4 for the different activation functions explored.

By analyzing the partial derivative $\frac{\partial \hat{Y}}{\partial S_o}$ the expected change in equity value $\hat{Y}$, per unit of change of the different scopes of emissions X_p, ceteris paribus, is shown. To be able to compare the effect of different input variables, standardized input data needs to be used. Using this method, potential (non-)linearity can be showed giving more insights into the hidden relation between equity value and the corporate carbon footprint as valued by investors.

4.8 Robustness

To check how robust the models are, this study performs several robustness checks. First, the models are evaluated using different metrics. Every metric has its own properties and can theoretically show different results. By comparing various metrics, specific behaviour of the models can be analyzed validating the robustness of the model. Second, several subsets of the data are used. In this way, it can be examined if the model also performs the same way in a specific region, industry or sector such as the energy and utility sector. The model where the corporate carbon footprint is estimated could serve as a general estimation method. If such a method would be implemented by policymakers it should perform relatively consistent across all segments, industries and regions.

5 Results

In this section, the best performing method for predicting the corporate carbon footprint is found. Second, the relation between equity value and the carbon footprint is explored.

5.1 Corporate carbon footprint

5.1.1 Prediction performance

Table 1 summarizes the out-of-sample prediction performance after hyperparameter tuning using the mean absolute error as tuning metric. The models are trained on 80% of the data and are tested on the remaining 20% using 5-fold-group-cross-validation. The metrics are calculated for every fold

and then averaged over these folds to increase robustness. Finally, all metrics are relative to the benchmark OLS.

Table 1: Prediction performance corporate carbon footprint estimation methods

Improvement	Scope 1			Scope 2			Scope 1+2		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
OLS	1.154	1.525	0.751	1.030	1.477	0.571	0.888	1.206	0.756
Linear methods									
Ridge	1.002	1.000	1.000	0.999**	1.000	1.000	1.000	1.000	1.000
Lasso	1.000	1.000	1.000	0.998**	0.999	1.001	1.000**	0.999	1.000
Elastic Net	1.000	0.999	1.000	0.999**	1.000	1.001	0.999**	0.999	1.001
Machine learning methods									
Random Forest	0.970***	0.970	1.020	0.930***	0.934	1.095	0.904***	0.924	1.048
LightGBM	0.930***	0.931	1.044	0.909***	0.919	1.116	0.872***	0.884	1.071
XGBoost	1.120**	1.090	0.938	0.990	0.972	1.041	1.022***	1.022	0.985
Neural Network	0.948***	0.940	1.039	0.925***	0.940	1.087	0.956***	0.940	1.038

*Note. The above panel shows the out-of-sample prediction performance using a 80/20% split of the data. The first row shows the out-of-sample MAE, RMSE and R^2 using 5-fold-group-cross-validation for the benchmark OLS. For the other methods, the relative improvement against the benchmark OLS is displayed by dividing the metric value of a specific method by the metric value of the benchmark. Here, a value < 1 for MAE and RMSE shows an improvement, a value > 1 for R^2 shows an improvement too. A value equal to 1 shows similar performance compared to the benchmark. The Diebold Mariano test is used to test the statistical significance of the improvement in MAEs. *, **, and *** represent statistical significance at 10%, 5% and 1% levels. Scope 1+2 represents the sum of scope 1 and scope 2 emissions as dependent variable.*

Table 1 shows the superior predictive performance of the machine learning methods. Machine learning outperforms the linear methods for all scopes which is in line with the results found in Nguyen et al. (2021). However, not all results are in line. Remarkably, XGBoost is performing worse than the linear models. Nguyen et al. (2021) found that XGBoost was the best performing predictor, however in this study only a small improvement for scope 2 emissions compared to the benchmark is found. A possible explanation could be that the hyperparameter tuning process resulted in different optimal values where a local optimum is found in this study where Nguyen et al. (2021) found a global one. Another explanation could be that the cleaning of the data set is slightly different resulting in different OOS performance. The small increase in average prediction performance of shrinkage methods is surprising too. Shrinkage showed good performance compared to a standard OLS in other studies (see Section 2) but the increase is small in this study. This can be explained by the relatively high number of observations compared to the number of predictors

resulting in little shrinkage. LightGBM shows the best predictive performance on all scopes. It outperforms the linear models and all machine learning models.

Looking at the variance of the individual errors in the sample indicated by a relatively large difference between the MAE and RMSE we see equal variance for all linear methods estimating all scopes. The variance increases slightly using machine learning methods indicating the benefit of adding a little bit of variance to increase model performance. Lastly, the R^2 is high for scope 1 and scope 1+2 for almost all machine learning methods, but highest for LightGBM. The R^2 is lowest for scope 2, indicating lower prediction performance for these emissions. Scope 2 emissions are indirect emissions that are more difficult to calculate and estimate. Since this model is trained on data disclosed by companies, the data for scope 2 emissions may be of lower quality resulting in lower prediction performance. Increasing data quality could lead to better performance in a future study. The combination of scope 1 and 2 emissions shows the best performance which can be explained by the combination of two estimations averaging out errors.

To be able to differentiate statistically between the methods the MCS is computed which includes the set of superior models. This set confirms the superior prediction performance of LightGBM since it solely includes LightGBM as superior model at a 5% significance level. This implies that regulators or investors can not choose a different method to increase for instance interpretability without losing significant prediction performance. See Appendix H for the results and p-values.

5.1.2 Predictor relations

The linear coefficients and SHAP values give insight in the influence of certain characteristics on the corporate carbon footprint.

Table 2: Predictor coefficients in linear models: corporate carbon footprint

Predictors	**OLS**			**Elastic Net**		
	Scope 1	Scope 2	Scope 1+2	Scope 1	Scope 2	Scope 1+2
Intercept	0.074	-2.401***	0.499***	-0.034	-2.437***	-0.115**
Log Revenue	0.496***	0.174***	0.256***	0.516***	0.195***	0.280***
Log EBITDA	0.040	0.162***	0.038	0.045	0.157***	0.063**
Log CapEx	0.423***	0.267***	0.357***	0.418***	0.269***	0.380***
Log PPE Age	0.196***	0.141***	0.152***	0.189***	0.146***	0.143***
Log PPE Net	0.351***	0.259***	0.387***	0.334***	0.253***	0.383***
Log Intangibles	-0.025*	0.001	-0.018*	-0.024	0.003	-0.018
Log Total Assets	0.134***	0.223***	0.262***	0.123***	0.209***	0.253***
Log Long Term Debt	0.091***	0.031	0.042**	0.092***	0.018	0.051**
Log FTE	0.309***	0.526***	0.401***	0.315***	0.543***	0.324***
Gross Margin	-0.225***	-0.071***	-0.141***	-0.230	-0.078***	-0.167***
Leverage (%)	0.042**	-0.019	0.019	0.040**	0.000	-0.003
Capital Intensity (%)	-0.034	-0.035**	-0.062***	0.000	-0.035**	-0.045***
Fuel Intensity (kgCO2-e/kWh)	0.117***	0.297***	0.254***	0.115***	0.305***	0.239***
Year dummy (baseline = 2007)						
2008	0.105	-0.049**	-0.045	0.194**	0.071	0.062
2009	0.020	-0.045**	-0.080	0.109	0.080	0.036
2010	-0.006	-0.097	-0.095	0.084	0.033	0.011
2011	-0.069	-0.138	-0.142*	0.000	0.000	-0.037
2012	-0.142	-0.187**	-0.202**	-0.053	-0.059	-0.098**
2013	-0.128	-0.203**	-0.208**	-0.039	-0.078*	-0.103**
2014	-0.117	-0.164*	-0.172**	0.000	-0.040	-0.061
2015	-0.040	-0.117	-0.109	0.048	0.000	0.006
2016	-0.095	-0.143	-0.132	0.000	0.000	0.000
2017	-0.190*	-0.247***	-0.227***	-0.101**	-0.123***	-0.116***
2018	-0.220**	-0.312	-0.258***	-0.131***	-0.187***	-0.151***
Industry dummy (base = Energy)						
Materials	-0.545***	1.557***	-0.231***	-0.530***	1.480***	0.214***
Capital Goods	-2.473***	0.078	-1.915***	-2.455***	0.000	-1.439***
Commercial & Professional Services	-2.612***	-0.389***	-2.020***	-2.599***	-0.477***	-1.500***
Transportation	-1.148***	-0.793***	-0.908***	-1.135***	-0.871***	0.000
Autombiles & Components	-3.019***	0.491***	-2.000***	-3.004***	0.406***	-1.566***
Consumer Durables & Apparel	-3.364***	-0.289***	-2.477***	-3.349***	-0.365***	-2.015***
Consumer Services	-2.534***	0.113	-1.678***	-2.517***	0.000	-1.150***
Retailing	-4.007***	-0.045	-2.508***	-3.987***	-0.138*	-2.008***
Food & Staples Retailing	-2.958***	0.437***	-1.909***	-2.947***	0.000	-1.426***
Food. Beverage & Tobacco	-1.672***	0.509***	-1.426***	-1.648***	0.425***	-0.940***
Household & Personal Products	-2.785***	-0.066	-2.245***	-2.760***	-0.144	-1.770***
Health Care Equipment & Services	-3.810***	-0.451***	-2.731***	-3.791***	-0.525***	-2.248***
Pharmaceuticals. Biotech & Life Sciences	-2.654***	-0.046	-2.180***	-2.626***	0.000	-1.710***
Diversified Financials	-5.021***	-1.333***	-3.538***	-4.988***	-1.363***	-3.071***
Insurance	-5.038***	-1.647***	-3.905***	-5.013***	-1.674***	-3.418***
Software & Services	-4.595***	-0.523***	-2.924***	-4.582***	-0.592***	-2.432***
Technology Hardware & Equipment	-4.074***	0.310***	-2.257***	-4.063***	0.227***	-1.777***
Semiconductors & Semiconductor Equipment	-3.088***	0.702***	-1.710***	-3.082***	0.634***	-1.260***
Telecommunication Services	-4.570***	0.351***	-2.413***	-4.563***	0.285***	-1.984***
Media & Entertainment	-4.269***	-0.427***	-2.863***	-4.252***	-0.490***	-2.407***
Utilities	0.086	0.790***	0.286***	0.094	0.736***	0.688***
Real Estate	-3.208***	0.808***	-1.635***	-3.195***	0.767***	-1.213***
CO2 Law dummy (base = National law)						
No CO2 Law implemented	-0.128***	0.026	-0.028	-0.120**	0.000	0.020
Other types of CO2 law	-0.029	0.421***	0.160***	-0.025	0.398***	0.197***
Regional implemented CO2 law	0.025	0.017	-0.124***	0.029	0.000	0.000
Sub-national implemented CO2 law	0.088	0.031	0.046	0.082	0.000	0.119**
Income Group Dummy (baseline = HI)						
Upper-Middle-income group	-0.121**	0.000	-0.027	-0.121**	0.000	-0.013
Low-Middle-income group	0.368	-0.262***	-0.002	0.369***	-0.287***	0.083

*Note. This table shows the roles of predictors across a selected number of methods. The coefficients are averaged coefficients from 5 folds. All continuous variables are standardized to get comparable partial correlations, dummy variables excepted. Scope 1+2 represents the sum of scope 1 and scope 2 emissions as dependent variable. *, **, and *** represent the statistical significance of the coefficients at 10%, 5% and 1% levels.*

Table 2 shows the roles of predictors in the OLS and Elastic Net regression. Being both linear methods, coefficients are easy to interpret and give a clear indication of the role of a predictor in the model. Most continuous predictors have a positive relation with the different scopes of the corporate carbon footprint. It is also notable that all significant continuous predictors have equal signs for the different scopes. Intangible assets are included in the total assets and do not have a carbon footprint themselves by definition. So if the proportion of intangible assets is large, it will decrease the corporate carbon footprint compared to the same company with fewer intangibles. The positive relation between leverage and scope 1 emissions is also found in Griffin et al. (2017). Another negative relation can be seen between the carbon footprint and the gross margin. This relation is most negative for scope 1 emissions which follows from the fact that a manufacturing company (low gross margin) needs to process raw materials, resulting in higher energy use and hence more scope 1 emissions compared to a service provider (high gross margin). Capital intense companies also tend to have a larger carbon footprint as is shown by the negative relation. Being more capital intensive can show that large and expensive projects are done by the company resulting in a higher carbon footprint. However, this is in contrast with findings in Goldhammer et al. (2016) where a positive sign is found. Nguyen et al. (2021) finds a high positive coefficient for the log power plant equipment net value (Log PPE Net). This study finds this positive relation too, however with a relatively smaller coefficient. Having expensive physical equipment possibly indicates a large machine park causal to a higher carbon footprint. Different industries and regions can have older equipment. As emissions become more important newer equipment tend to have lower emissions. The positive coefficient for the age of power, plant and equipment underlies this argument implying higher emissions for older equipment.

Next, the continuous coefficient values are compared for the different scopes. Since the scopes have different types of emissions, different predictors can be important. Revenue for instance has a higher positive relation with scope 1 emissions than with scope 2 emissions since its effect is shared with other predictors that are related to the scale of operations. The coefficient for the number of employees (FTE) is higher for scope 2 than for scope 1 as is the coefficient for total assets. This is evident since scope 2 emissions follow directly from the carbon footprint of employees and the heating or cooling of offices.

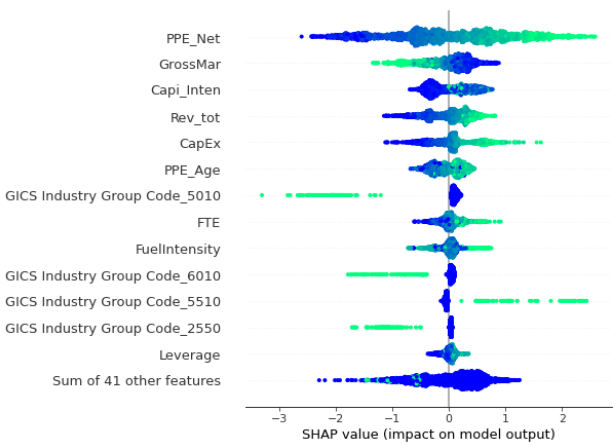
First, when looking at the coefficients for the year dummies a negative trend compared to the baseline year 2007 is present for most significant coefficients. This was to be expected since companies are pressured to reduce their carbon corporate footprint. The elastic net, however, shows positive

and negative relations for all scopes estimated but almost all positive coefficients are statistically insignificant. Secondly, the baseline for the industry sector is the energy sector. The GICS industry sectors are used to distinguish 24 different industries. All significant scope 1 coefficients are negative indicating the highest scope 1 emissions of all industries for the energy industry. Other industries with a large scope 1 footprint are the utilities, materials and transportation industries. This follows logically from the definition of their activities and the associated energy consumption. The energy industry does not necessarily have the largest scope 2 emissions since they cover indirect emissions. This explains the diversity in signs and magnitudes of coefficients for scope 2 emissions. Combining scope 1 and 2 emissions show that the energy, materials and transportation sector are the biggest polluters.

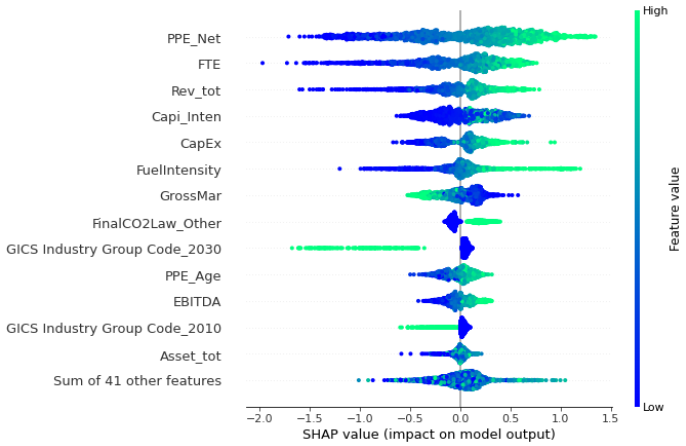
Thirdly, the impact of the existence of a CO_2 law is looked into. Having a national CO_2 law is set as a baseline. The linear models show inconsistent effects off different types of CO_2 laws on the scopes of emissions. Against expectations, having no CO_2 law implemented shows lower scope 1 emissions. Compared to the national law, having a sub-national implemented CO_2 law is counter-effective to decrease the carbon footprint for scope 1+2 emissions. A national CO_2 law results in lowest scope 2 emissions. Implementing a regional CO_2 law results in lowest scope 1+2 emissions.

Finally, the income level of the country headquarters is looked at. The high-income group is set as a baseline. As expected, countries with low middle income show higher emissions since there will be less focus on reducing the carbon footprint. Interestingly, upper-middle-income countries show lower scope 1 and 1+2 emissions possibly indicating that richer countries tend to emit more Scope 1 emissions.

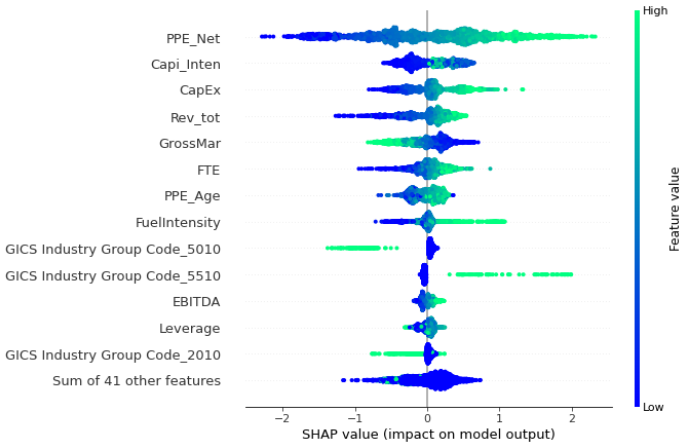
Figure 3: SHAP Beeswarm plot predictor importance estimation corporate carbon footprint.



(a) Predictor importance Scope 1.



(b) Predictor importance Scope 2.



(c) Predictor importance Scope 1+2.

Note. The above figures show the summary beeswarm SHAP plots for each estimated scope. The scopes are estimated using LightGBM. The different predictors are listed on the left side in decreasing order of importance. Each point of every row is an observation of the test data set and the x-position is determined by the corresponding SHAP value. Colour is used to display the original value of the feature. See Appendix D for the industry classification codes.

Section 5.1.1 shows that LightGBM outperforms the other methods consistently. SHAP values are calculated for the LightGBM method and the beeswarm SHAP summary plots are shown in Figure 3. They show the predictor importance of the 13 most important predictors used by LightGBM. The plot is designed to display a summary of how the top features in a data set impact the output of the model. The rows consist of many dots where each dot represents one observation. The horizontal position of the dot is determined by the SHAP value of that predictor for that observa-

tion. The colour scale on the right of the figure is used to show the original value of the feature. The predictors are ordered using the mean absolute value of the SHAP values for each feature to place more emphasis on broad average impact, and less on rare but high magnitude impacts. This is of more interest in this study since predictors that are important for a general estimation method are tried to be found rather than predictors with high individual impact. The horizontal location of a dot shows whether the effect of that observation results in a higher or lower prediction of the dependent variable.

The figures show that the net value of property, plant and equipment (PPE_net) is the most important predictor on average for the estimation of all scopes. The colour shows that a low value for PPE_net results in a negative and high impact on the prediction of all scopes. Other important predictors for scope 1 are the gross margin (GrossMar), capital intensity (Capi_inten) and total revenue (Rev_tot). These predictors all represent the scale of operations showing its importance to the size of the corporate carbon footprint. Gross margin has a negative effect on the dependent variable, which is in line with the results found in Section 5.2.2. Interestingly, only industry group dummies show some importance in the estimation of all scopes, where year and CO_2 law dummies also showed relatively large coefficients in the linear models. The utility industry (code_5510) is as expected positively correlated with the scope 1 emissions. The dots are concentrated to the positive side of the figure indicating almost only positive impacts on the dependent variable. The real-estate industry (code_6010), the retailing industry (code_2550 and the telecommunication industry (code_5010) show lower scope 1 emissions than the energy sector that served as baseline. The amount of full-time employees (FTE) has a high and positive impact on the scope 2 emissions as was found in the linear models. For scope 2 emissions fewer dummy variables are important here. The dummy that indicates other types of CO_2 laws has a small positive impact on the scope 2 emissions as was found in the linear models. Scope 1+2 emissions show a similar pattern as found in scope 1 emissions. It is interesting to see that the separate year dummies and their joint impact are not important for the estimation of all scopes. The joint importance has a SHAP value of 0.018 where the least important predictor in Figure 3 has an average SHAP value of 0.083 for scope 1 emissions.

5.2 Equity value

5.2.1 Prediction performance

In order to predict equity value, scope 1, 2 and 1+2 emissions are used as separate predictor variables. Table 3 summarizes the out-of-sample prediction performance after hyperparameter tuning using the mean absolute error as tuning metric. Since the focus does not lay on the comparison of models but rather on the relations within a model, a brief performance review is given in this section.

Table 3: Prediction performance equity value

Metric	Scope 1			Scope 2			Scope 1+2		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
OLS	0.412	0.539	0.839	0.414	0.540	0.838	0.413	0.540	0.838
Linear methods									
Ridge	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Lasso	1.002	1.001	1.000	0.999	1.000	1.000	0.999	0.999	1.000
Elastic Net	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Machine learning methods									
Random Forest	0.932***	0.932	1.025	0.940***	0.939	1.023	0.929***	0.930	1.026
LightGBM	0.918***	0.927	1.027	0.921***	0.930	1.026	0.918***	0.928	1.027
XGBoost	0.980*	0.968	1.012	1.103	1.072	0.971	1.102	1.071	0.972
Neural Network	0.839***	0.853	1.038	0.836***	0.841	1.042	0.831***	0.846	1.040

*Note. The above panel shows the out-of-sample prediction performance using an 80/20% split of the data. The first row shows the out-of-sample MAE, RMSE and R^2 using 5-fold-group-cross-validation for the benchmark OLS. For the other methods, the relative improvement against the benchmark OLS is displayed by dividing the metric value of a specific method by the metric value of the benchmark. Here, a value < 1 for MAE and RMSE shows an improvement, a value > 1 for R^2 shows an improvement too. A value equal to 1 shows similar performance compared to the benchmark. The Diebold Mariano test is used to test the statistical significance of the improvement in MAEs. *, **, and *** represent statistical significance at 10%, 5% and 1% levels.*

The results from Table 3 show similar results as for the first model. Again, machine learning appears to enhance prediction performance. Only XGBoost underperforms which could be explained by subpar hyperparameter tuning or data cleaning. The best estimator in the previous section, LightGBM, shows again better prediction performance than the benchmark. It is the second-best performing algorithm after the neural network. The neural network shows excellent prediction performance compared to the other methods and especially the linear methods. With a R^2 of 87,3% the amount of variance explained is high. The reduction of both MAE and RSME is high too with relative decreases of over 10% for all scopes. Superior prediction performance is also underlined by

the MCS which includes the Neural Network in its superior model set using a 5% significance level. See Appendix H for the the different p-values.

5.2.2 Predictor relations

The potential hidden costs of the carbon footprint as valued by investors can be shown using the coefficients in the linear models. SHAP values will not be able to quantify such a relation directly, but it can show if certain scopes of emissions play an important role in the prediction of the equity value. Furthermore, it is able to show if this relation is positive or negative. First-order partial derivatives are used to show potential (non-)linearity.

Table 4: Predictor coefficients in linear models: log equity value

Predictors	**OLS**			**Elastic Net**		
	Scope 1	Scope 2	Scope 1+2	Scope 1	Scope 2	Scope 1+2
Intercept	22.769***	22.711***	22.734***	22.766***	22.731***	22.730***
Scope X emissions	-0.053***	0.045***	-0.018**	-0.053***	0.046***	-0.018**
Log Total Assets	0.514***	0.515***	0.513***	0.515***	0.515***	0.513***
Log Book Value	0.271***	0.252***	0.266***	0.272***	0.252***	0.267***
Log Income Net before taxes	0.265***	0.268***	0.267***	0.266***	0.270***	0.268***
Log Total Liabilities	-0.303***	-0.334***	-0.311***	-0.304***	-0.334***	-0.312***
Log Operating Income	0.524***	0.517***	0.521***	0.523***	0.513***	0.520***
CO2 Law dummy (base= No law)						
Subnational implemented CO2 law	0.134***	0.154***	0.135***	0.137***	0.155***	0.138***
National implemented CO2 law	0.255***	0.252***	0.254***	0.257***	0.254***	0.257***
Regional implemented CO2 law	-0.013	-0.007	-0.015	0.000	0.000	0.000
Other CO2 law	0.106***	0.113***	0.105***	0.109***	0.112***	0.108***
Year dummy (baseline = 2007)						
2008	-0.562***	-0.565***	-0.565***	-0.562***	-0.565***	-0.565***
2009	-0.107***	-0.106***	-0.109***	-0.106***	-0.106***	-0.108***
2010	-0.123***	-0.120***	-0.124***	-0.123***	-0.119***	-0.123***
2011	-0.357***	-0.352***	-0.357***	-0.357***	-0.351***	-0.356***
2012	-0.233***	-0.224***	-0.231***	-0.232***	-0.223***	-0.230***
2013	-0.078**	-0.073*	-0.077**	-0.077	-0.072*	-0.077**
2014	-0.126***	-0.123***	-0.125***	-0.125***	-0.122***	-0.125***
2015	-0.164***	-0.163***	-0.165***	-0.164***	-0.161***	-0.164***
2016	-0.152***	-0.147***	-0.151***	-0.152***	-0.146***	-0.150***
2017	-0.093***	-0.084**	-0.091***	-0.093***	-0.083**	-0.091**
2018	-0.335***	-0.325***	-0.333***	-0.335***	-0.324***	-0.333***
Industry dummy (base = Energy sector)						
Materials	0.001	-0.037	0.003	0.002	-0.060***	0.004
Capital Goods	0.010	0.048*	0.038	0.010	0.026	0.039
Commercial & Professional Services	0.124***	0.179***	0.156***	0.125***	0.156***	0.157***
Transportation	0.031	0.056*	0.043	0.031	0.034	0.044
Automobiles & Components	-0.214***	-0.190***	-0.182***	-0.215***	-0.213***	-0.183***
Consumer Durables & Apparel	0.025***	0.097***	0.067*	0.026	0.075**	0.068*
Consumer Services	0.254***	0.272***	0.279***	0.254***	0.250***	0.280***
Retailing	0.103***	0.157***	0.151***	0.103***	0.135***	0.152***
Food & Staples Retailing	0.222***	0.232***	0.250***	0.224***	0.210***	0.252***
Food. Beverage & Tobacco	0.280***	0.291***	0.297***	0.281***	0.269***	0.299***
Household & Personal Products	0.601***	0.638***	0.632***	0.601***	0.615***	0.631***
Health Care Equipment & Services	0.350***	0.431***	0.399***	0.350***	0.408***	0.399***
Pharmaceuticals. Biotech & Life Sciences	0.440***	0.502***	0.476***	0.440***	0.479***	0.476***
Diversified Financials	-0.233***	-0.040	-0.144***	-0.232***	-0.062***	-0.142***
Insurance	-0.467***	-0.267***	-0.382***	-0.466***	-0.288***	-0.380***
Software & Services	0.399***	0.502***	0.466***	0.400***	0.479***	0.467***
Technology Hardware & Equipment	-0.011	0.050	0.043	-0.010	0.000	0.045
Semiconductors & - Equipment	0.136***	0.167***	0.177***	0.136***	0.144***	0.177***
Telecommunication Services	0.066*	0.137***	0.129***	0.066*	0.115***	0.129***
Media & Entertainment	0.079*	0.194***	0.144***	0.077*	0.170***	0.143***
Utilities	-0.194***	-0.190***	-0.190***	-0.195***	-0.213***	-0.191***
Real Estate	-0.309***	-0.194***	-0.243***	-0.309***	-0.216***	-0.241***

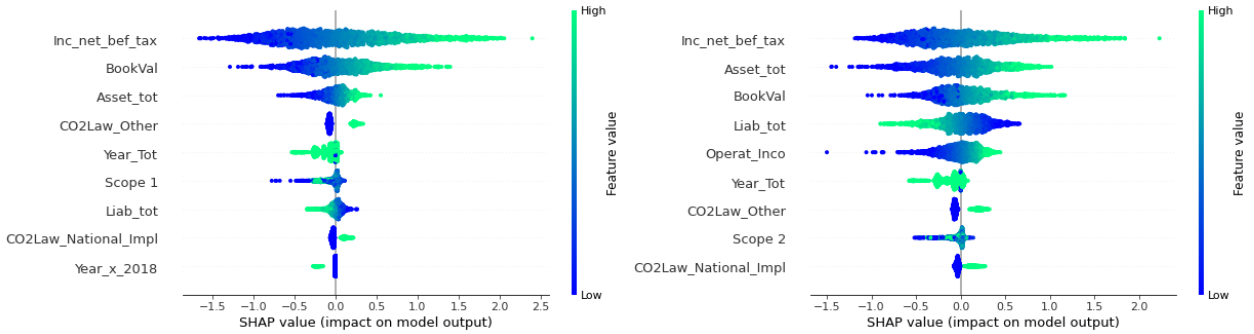
*Note. This table shows the roles of predictors across a selected number of methods. The coefficients are averaged coefficients from 5 folds group-cross-validation. All continuous variables are standardized to get comparable partial correlations, dummy variables excepted. The dependent variable and specific continuous predictors are log transformed (see 3.4). Here, the columns indicate which scope of emissions is included as predictor variable. The column scope 1+2 means for instance that the sum of scope 1 and scope 2 emissions are included as predictor. *, **, and *** represent the statistical significance of the coefficients at 10%, 5% and 1% levels.*

First, the most interesting continuous variable is explored, namely the relation between the different scopes of emissions and equity value. This relation represents the potential hidden costs in the corporate carbon footprint as it is valued by investors. Since investors are pressured by regulators and the public to take ESG scores into consideration a negative relation is to be expected as is found by Chapple et al. (2013), Matsumura et al. (2014) and Clarkson et al. (2015). Looking at scope 1 and scope 1+2 emissions in Table 4, this relation is found. Since the dependent and the continuous predictor variables are both log-transformed, the coefficient can be seen as the per cent decrease in the dependent variable for every 1% increase in the predictor variable if the coefficient is negative. So we see, that if scope 1 emissions increase by 1% the equity value of that same company decreases by 0.053%. Taking the scope 1+2 coefficient, it would mean a decrease of 0.018%. Interestingly, scope 2 emissions tend to have a positive effect on equity value. This could be caused by a larger focus on scope 1 (direct) emissions instead of scope 2 (indirect) emissions.

Table 4 displays the predictor relations for the equity model. All continuous variables have the expected sign. It is evident that the total assets, book value, net income before taxes and operating income all have a positive relation with equity value. The fact that the size of the coefficient stays relatively constant when using different scopes as predictor variable shows the robustness of these continuous variables. Operating income and total assets show the largest coefficients. If a company has a high amount of total assets its equity value will be large too. High operating income also shows high equity value since the company is profitable.

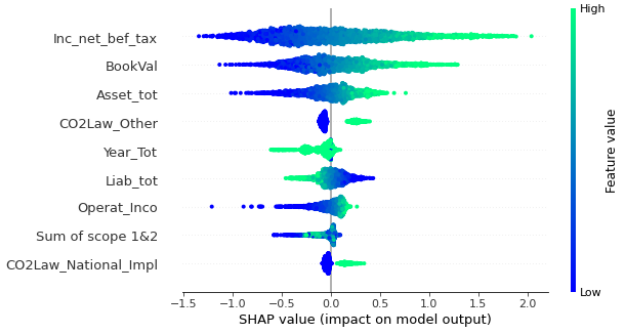
Next, the dummy variables are explored. 2007 serves as a baseline for the yearly dummy and the other coefficients are compared to this baseline. All coefficients are negative which can be explained by the economic crisis of 2008 where equity value decreased drastically. Note that this year has the most negative coefficient too. Next, the industry dummy is analyzed. The household & personal products industry has the highest coefficient in relation to equity value, where real-estate gets the lowest. More interesting is the CO_2 law dummy indicating the impact of a CO_2 law on the equity value of companies that have their headquarters in that specific area. The baseline is no CO_2 law and the other options are compared against this baseline. The positive coefficients for a (sub-)national and another type of CO_2 law imply that the introduction of a CO_2 law has a positive impact on the equity value of companies. This is contrary to arguments of lobbyists against the implementation of a CO_2 law that state that this would have a negative effect on the companies listed in the specific country. Insignificant negative coefficients for the implementation of a regional CO_2 law are found.

Figure 4: SHAP Beeswarm plot predictor importance equity value.



(a) Predictor importance using Scope 1 as predictor

(b) Predictor importance using Scope 2 as predictor



(c) Predictor importance using Scope 1+2 as predictor

Note. The above figures show the summary beeswarm SHAP plots estimating log equity value using different scopes of emissions included in the predictor set. The scopes are estimated using Neural Networks. The different predictors are listed on the left side in decreasing order of importance. Each point of every row is an observation of the test data set and the x-position is determined by the SHAP value. Colour is used to display the original value of the feature. See Appendix D for the industry classification codes.

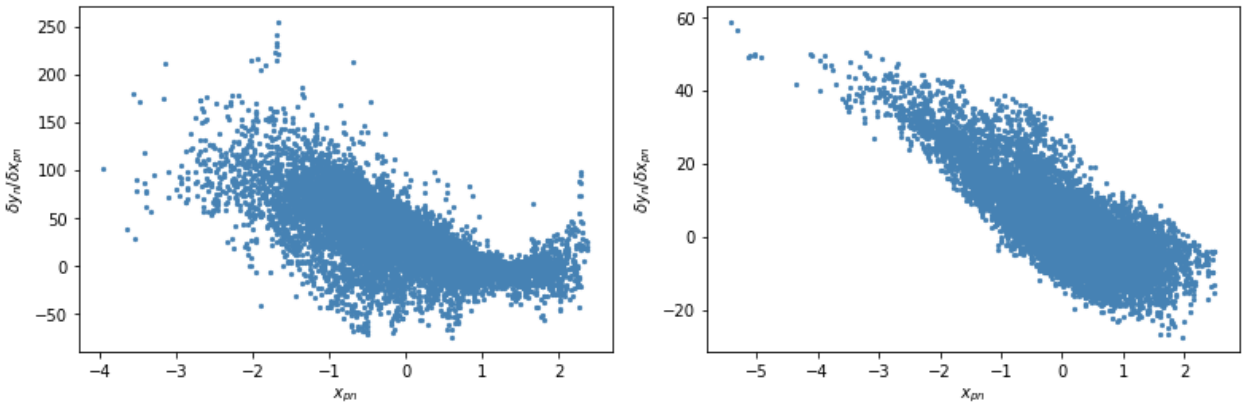
To be able to say something about the predictor relations in the neural network, SHAP values are computed via the method described in Section 4.6 and are displayed using beeswarm SHAP plots shown in Figure 4. The most important predictors listed on the left side of the figures in decreasing order. The figures show that the net income before taxes (Inc_net_bef_tax) has the highest positive impact on the equity value in all three models which is in line with the linear models discussed in Section 5.2.2. Other predictors with high impact are the book value (BookVal), total assets (Asset_tot) and the different scopes of emissions. The positive impact of the implementation of national or other types of CO_2 law is also found in the neural network. The separate year dummies are of low importance, however, the joint importance of the year dummy is relatively high. The

negative relation is in line with the results from the linear models.

It is interesting to see that overall the scopes of emissions have a negative impact on the equity value in general indicating the hidden costs as valued by investors for the corporate carbon footprint. However, the colour scales implies that higher than average amounts of emissions result in near-zero or a small positive impact on the equity value which was not expected. This could imply that there is some kind of threshold on emissions. Under this threshold, carbon emissions are negatively related to equity value, however, whenever this threshold is surpassed the emissions are ignored by investors eliminating the negative relation with equity value.

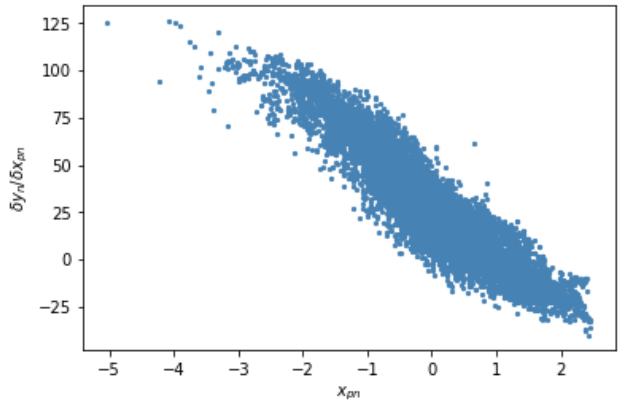
To explore such a non-linear relation a sensitivity analysis is performed. Figure 5 shows the partial derivatives of the output y_n with respect to the inputs x_{pn} in the neural network models that estimate equity value using scope 1, 2 and 1+2 emissions respectively ($p = 1$). Panel a shows a highly non-linear relation. It also confirms the near-zero or slightly positive relation for higher Scope 1 emissions indicating the existence of a threshold found using SHAP values. The graphs for scope 2 and 1+2 shows a relation that somewhat looks non-linear since the negative effects flatten slightly for larger footprints. The relation also flattens for lower scope 2 and 1+2 emissions. This non-linearity shows that investors value different sizes of the corporate carbon footprint differently. If policymakers decide to penalize emissions equally, heavy polluters would be impacted the most, since a less negative relation is found now.

Figure 5: Partial derivatives of the output y_n with respect to inputs x_{pn}.



(a) Partial derivatives using scope 1 as predictor.

(b) Partial derivatives using scope 2 as predictor.



(c) Partial derivatives using scope 1+2 as predictor.

Note. The above figures show the partial derivatives of the output y_n representing the log equity value with respect to the inputs x_{pn} in the optimal neural network models that estimate equity value using scope 1, 2 and 1+2 emissions respectively ($p = 1$). The partial derivatives are showed for all n observations in the training set.

5.3 Robustness

To check the robustness of the models, several metrics are used as described in Section 4.5. Since these metrics use several ways of displaying the estimation error it gives a good insight into the robustness of the model. All metrics result in the same ranking of prediction performance for both estimated models. Some disparity is noted in these improvements comparing different metrics for ridge and Lasso regression. Ridge shrinkage results for instance in a 0.15% increase in MAE, however looking at the RMSE a 0.02% decrease is noted showing inconsistent results. However, since the

results are centred around zero, this lack of robustness is ignored.

As a second robustness test, a different subset of the data is used to train the models. This subset excludes companies that did not disclose their scope 3 emissions. The remaining subset has 8515 observations and 1696 companies. This is a 36% decrease in the number of companies compared to the original data set limiting the diversity of companies to learn from. Using this set, LightGBM is again the best performing method to predict the corporate carbon footprint. However, the different metrics show slightly inconsistent results. For instance, the R^2 using LightGBM decreases for scope 1 emissions from 78.4% to 76.1% using the smaller data set. Next, the second model where equity value is estimated is analyzed. Again a smaller data set without companies that did not disclose their scope 3 emissions is used. In this data set, 1611 companies are included with 7561 yearly-company observations. Using this data, the metrics show slightly different results. According to the R^2 the best performing algorithm is LightGBM (2.08% increase in R^2 for using scope 1 against 1.48% increase for the neural network). However, according to the MAE and RMSE the best performing algorithm is the neural network. Since the MAE and RMSE give a more thorough indication of the predictive performance of a method, the neural network is still preferred over LightGBM.

6 Conclusion and discussion

This study analyzes the predictability of the corporate carbon footprint concerning scope 1, 2 and 1+2 emissions. Furthermore, the hidden costs of the corporate carbon footprint as valued by investors are analysed estimating the equity value of companies using, amongst others, the carbon footprint as predictor. This analysis is performed on companies across all regions, sectors and industries to get an all-encompassing result. Different linear models, ensemble methods and neural networks are trained on publicly available data from the period 2007-2018. The robustness of the study is validated using 5-fold-group-cross-validation, different performance metrics and two subsets of data.

As the first main finding, this study shows the superior prediction performance of machine learning compared to the benchmark OLS in predicting the corporate carbon footprint. Until early 2021, linear models were solely used to estimate the carbon footprint. The random forest model improves the prediction performance already, but it is outperformed by the neural network. This study is the first to use LightGBM as an estimation method and it is found to be the best prediction method for the corporate carbon footprint. The superior results of machine learning are in line

with the findings in Nguyen et al. (2021). However, they find Extreme Gradient Boosting as a superior prediction method which this study does not. A theoretical explanation could be a different data cleaning method or hyperparameter tuning process. This can result in convergence to a local optimum instead of a global one. Concluding, greenhouse gas emissions can be best estimated using a specific set of publicly available predictor variables described in this study using the LightGBM estimation method.

To be able to create economical insights from the estimation models, the predictor importance is analyzed by looking at coefficients of linear models, SHAP values for the best performing method and first-order partial derivatives. Predictors related to the size of operations have the highest positive impact on the carbon footprint. The utilities, energy, materials and transport industries are identified as heavy polluters. A negative impact of the gross margin is found indicating that a company early in the supply chain uses more energy to process raw materials, with high levels of scope 1 emissions as result. The sign of the effect, however, is in contrast with Goldhammer et al. (2016). This study also finds that the age of the property, plant and equipment is positively related to the carbon footprint endorsing the fact that equipment has become less polluting over the years. By including a yearly dummy it is shown that all scopes of emissions are declining as compared to the baseline 2007. Surprisingly, having no CO_2 law implemented results in lower scope 1 emissions as compared to having a national law implemented. Also having a sub-national or regionally implemented CO_2 law is counter-effective to decrease the scope 1+2 emissions. However, the effects of different types of CO_2 law are inconsistent across the scopes. As expected, countries with low-middle-income show higher scope 1 and 1+2 emissions than high-income countries, however, upper-middle-income countries show lower scope 1 emissions.

Secondly, the hidden costs of the corporate carbon footprint as valued by investors is explored. Again, linear models, ensemble methods and neural networks are used to explore the contemporaneous relation between equity value and a specific scope of emission. Company data that consists of observations across regions, sectors and industries are used. Machine learning methods show superior predictive performance. In contrary to the first model, LightGBM was the second to best performer. Neural networks with 1 and 2 hidden layers showed superior predictive performance. All models have high R^2 ranging from 83% to 87%. This study uses the corporate carbon footprint as a predictor to This study uses the corporate carbon footprint as a predictor to explore the contemporaneous relation with equity value.

To obtain insights into the hidden costs of the corporate carbon footprint the predictor relations

from the second model are analyzed. The linear models show a negative coefficient for the scope 1 and scope 1+2 emissions which is in line with findings in Matsumura et al. (2014) and Griffin et al. (2017). Surprisingly, it shows a positive coefficient for scope 2 emissions. OLS finds a price elasticity of -0.053% for scope 1 emissions and -0.018% for scope 1+2 emissions representing the negative valuation effect as valued by investors. Interestingly, implementing a (sub-)national or another type of CO_2 law has a positive effect on the equity value. This endorses the hypothesis that potential opportunities coming from climate change-related laws outweigh the potential costs.

SHAP values show the highest impact for business-related predictor variables and different scopes of emissions. It finds negative relations between equity value and the carbon footprint in the best performing neural network. However, looking at observations with highly above average amounts of emissions a near-zero or small positive impact on the equity value is found. The performed sensitivity analysis using first-order partial derivatives underlines this conclusion since highly non-linear relations are found for scope 1 emissions and slightly non-linear relations for scope 2 and 1+2 emissions. It shows that heavy polluting companies get a lower negative discount on their equity value. This could imply that there is some kind of threshold on emissions. Under this threshold, a negative relation is found, however, when the threshold is exceeded the emissions are ignored by investors eliminating the negative relation with equity value. If policymakers decide to penalize emissions equally, heavy polluters would be impacted the most, since a less negative relation is found now.

The conclusions that follow from this study are deemed valid according to several robustness checks. However, the lack of performance of the Extreme Gradient Boosting algorithm is surprising. A more extensive hyperparameter tuning procedure should be performed in order to converge to the global optimum.

Another point of discussion is the hyperparameter tuning overall. Ideally, a gridsearch over all hyperparameters is performed testing every single combination to find the optimal model. Since this is computationally too expensive a selection of hyperparameters has been made. Some hyperparameters are sequentially trained too which can also result in convergence to a local optimum.

Most interestingly is further research on the non-linearity and the potential threshold. It is interesting to see how the negative relation between equity value and the carbon footprint behaves in different sectors, industries and regions. Secondly, further research could look into the surprising impact of different types of CO_2 laws. It is interesting to see how the implementation of a CO_2 law increases the corporate carbon footprint in some cases.

References

H. Alshari, A. Saleh, and A. Odaba. Comparison of gradient boosting decision tree algorithms for CPU performance. *Journal of Institue of Science and Technology*, 37(1):157–168, 2021.

S. Alvarez, M. Blanquer, and A Rubio. Carbon footprint using the compound method based on financial accounts. The case of the school of forestry engineering, technical university of Madrid. *Journal of Cleaner Production*, 66:224–232, 2014.

C. M. Bishop. *Neural networks for pattern recognition.* Oxford University Press, 1995.

T. Busch, M. Johnson, and T. Pioch. Corporate carbon performance data: Quo vadis? *Journal of Industrial Ecology*, 2020.

J. Cagiao, B. Gómez, J. L. Doménech, S. G. Mainar, and H. G. Lanza. Calculation of the corporate carbon footprint of the cement industry by the application of MC3 methodology. *Ecological Indicators*, 11(6):1526–1540, 2011.

A. Carballo-Penela and J. L. Doménech. Managing the carbon footprint of products: The contribution of the method composed of financial statements (MC3). *The International Journal of Life Cycle Assessment*, 15(9):962–969, 2010.

L. Chapple, P. Clarkson, and D. Gold. The cost of carbon: Capital market effects of the proposed emission trading scheme (ETS). *Abacus*, 49(1):1–33, 2013.

T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM Sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

P. Clarkson, X. Fang, Y. Li, and G. Richardson. The relevance of environmental disclosures for investors and other stakeholder groups: Are such disclosures incrementally informative? *Journal of Accounting and Public Policy*, 32(5):1–33, 2013.

P. M. Clarkson, Y. Li, M. Pinnuck, and G. D. Richardson. The valuation relevance of greenhouse gas emissions under the European Union carbon emissions trading scheme. *European Accounting Review*, 24(3):551–580, 2015.

S. Eskander and S. Fankhauser. Reduction in greenhouse gas emissions from national climate legislation. *Nature Climate Change*, 10:750–756, 2020.

J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of statistics*, 5:1189–1232, 2001.

J. Friedrich, M. Ge, and A. Pickens. Interactive chart shows changes in the world's top 10 emitters, 2020. URL `https://www.wri.org/blog/2020/12/interactive-chart-top-emitters`.

G. Giese, L. E. Lee, D. Melas, Z. Nagy, and L. Nishikawa. Foundations of ESG investing: How ESG affects equity valuation, risk, and performance. *The Journal of Portfolio Management*, 45 (5):69–83, 2019.

B. Goldhammer, C. Busse, and T. Busch. Estimating corporate carbon footprints with externally available data. *Journal of Industrial Ecology*, 21(5):1165–1179, 2016.

P. Griffin, D. Lont, and E. Y. Sun. The relevance to investors of greenhouse gas emission disclosures. *Contemporary Accounting Research*, 34(2):1265–1297, 2017.

I. H. V. Gue, A. T. Ubando, M. L. Tseng, and T. R. Tan. Artificial neural networks for sustainable development: A critical review. *Clean Technologies and Environmental Policy*, pages 1–17, 2020.

P. R. Hansen, A. Lunde, and J. M. Nason. The model confidence set. *Econometrica*, 79(2):453–497, 2011.

V. H. Ho. Non-financial reporting & corporate governance: Explaining American divergence & its implications for disclosure reform. *Accounting, Economics, and Law: A Convivium*, 10(2), 2020.

A. E Hoerl and R. W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

World Research Institute and World Business Council for Sustainable Development. The greenhouse gas protocol. a corporate accounting and reporting standard. *World Business Council for Sustainable Development and World Resources Institute: Geneva, Switzerland*, page 116, 2004.

I. Ioannou and G. Serafeim. The consequences of mandatory corporate sustainability reporting. *Harvard Business School research working paper*, (11-100), 2017.

P. Kadam and S. Vijayumar. Prediction model: CO2 emission using machine learning. In *2018 3rd International Conference for Convergence in Technology (I2CT)*, pages 1–3. IEEE, 2018.

G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Y. Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30: 3146–3154, 2017.

D. P Kingma and J. Ba. Adam: A method for stochastic optimization. In *3rd International Conference for Learning Representations*, 2014.

A. Kolk, D. Levy, and J. Pinkse. Corporate responses in an emerging climate regime: The institutionalization and commensuration of carbon disclosure. *European Accounting Review*, 17(4): 719–745, 2008.

W. Leontief. Environmental repercussions and the economic structure: an input-output approach. *The Review of Economics and Statistics*, pages 262–271, 1970.

M. Lu, S. M. AbouRizk, and U. H. Hermann. Sensitivity analysis of neural networks in spool fabrication productivity studies. *Journal of Computing in Civil Engineering*, 15(4):299–308, 2001.

S. M Lundberg and S. I. Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems*, pages 4768–4777, 2017.

X. Ma, J. Sha, D. Wang, Y. Yu, Q. Yang, and X. Niu. Study on a prediction of P2P network loan default based on the machine learning lightgbm and xgboost algorithms according to different high dimensional data cleaning. *Electronic Commerce Research and Applications*, 31:24–39, 2018.

E. Matsumura, R. Prakash, and S. Vera-Muñoz. Firm-value effects of carbon emissions and carbon disclosures. *The Accounting Revieb*, 89(2):675–724, 2014.

W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133, 1943.

T. Mitton. Methodological variation in empirical corporate finance. *Available at SSRN 3304875*, 2020.

M. Nachmany, S. Fankhauser, J. Setzer, and A. Averchenkova. Global trends in climate change legislation and litigation: 2017 update. *Grantham Research Institute on Climate Change and the Environment*, 2017.

A. C Ng and Z. Rezaee. Business sustainability performance and cost of equity capital. *Journal of Corporate Finance*, 34:128–149, 2015.

Q. Nguyen, I. Diaz-Rainey, and D. Kuruppuarachchi. Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach. *Energy Economics*, 95(1):105–129, 2021.

V. Nourani and M. S. Fard. Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes. *Advances in Engineering Software*, 47(1):127–146, 2012.

M. Paliwal and U. A. Kumar. Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, 36(1):2–17, 2009.

D. Pandey, M. Agrawal, and J. S. Pandey. Carbon footprint: current methods of estimation. *Environmental Monitoring and Assessment*, 178:135–160, 2011.

J. Poushter and C. Huang. Climate change still seen as the top global threat, but cyberattacks a rising concern. *Pew Research Center*, 10(1):1–37, 2019.

Carbon Disclosure Project. Major risk or rosy opportunity: Are companies ready for climate change? URL `https://www.cdp.net/en/research/global-reports/global-climate-change-report-2018/climate-report-risks-and-opportunities`. Accessed 14 Apr. 2021.

P. Ramachandran, B. Zoph, and Q. V. Le. Searching for activation functions. *arXiv:1710.05941*, 2017.

M. Re and G. Valentini. Ensemble methods. *Advances in Machine Learning and Data Mining for Astronomy*, pages 563–593, 2012.

Y. Ren, L. Zhang, and P. N. Suganthan. Ensemble classification and regression-recent developments, applications and future directions. *IEEE Computational intelligence magazine*, 11(1):41–53, 2016.

C. Saleh, L. R. A. Chairdino, R. M. N. Ab, D. B. Md, and N. R. Dzakiyullah. Prediction of CO2 emissions using an artificial neural network: The case of the sugar industry. *Advanced Science Letters*, 21(10):3079–3083, 2015.

H. L. Shang and S. Haberman. Model confidence sets and forecast combination: An application to age-specific mortality. *Genus*, 74(1):1–23, 2018.

L. S. Shapley. *A value for n-person games*. Princeton University Press, 2016.

P. Werbos. Applications of advances in nonlinear sensitivity analysis. *System Modeling and Optimization*, pages 762–770, 1982.

T. Wiedmann. *Carbon footprint and input–output analysis–an introduction*. Taylor & Francis, 2009.

T. Wiedmann and J. Minx. A definition of 'carbon footprint'. *Journal of Political Economy*, 1:1–11, 2008.

Worldbank. State and trends of carbon pricing 2017. URL `http://documents.worldbank.org/`. Accessed 07 Apr. 2021.

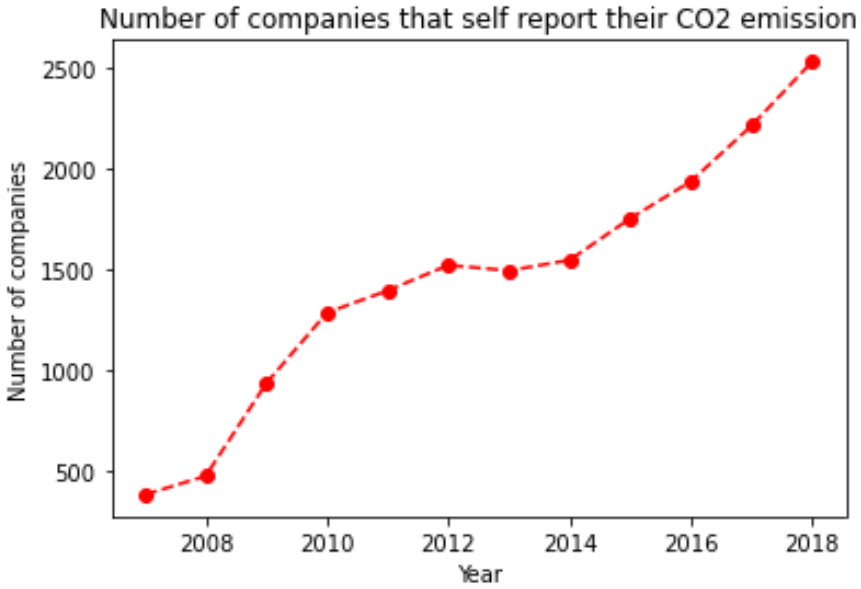
G. Xu, P. Schwarz, and H. Yang. Determining China's CO2 emissions peak with a dynamic nonlinear artificial neural network approach and scenario analysis. *Energy Policy*, 128:752–762, 2019.

Appendices

A Companies that disclose their CO_2 emission over time

In Figure 6, the number of companies that disclose their own emissions is displayed over the period 2007-2018.

Figure 6: Number of disclosers over time



Note. The data used consists only of companies that disclose its own emissions. In the graph it is shown that the number of companies that disclose their emissions is growing rapidly.

B Carbon footprint calculation methods

Companies can use different methods to disclose their corporate carbon footprint. The process-product-based life-cycle assessment (P-LCA) is based on ISO standards. Unfortunately, the method is not applicable to all sectors. According to Cagiao et al. (2011), the method has a number of problems especially in terms of comparability. P-LCA uses a bottom-up approach taking a product as a starting point. It then calculates all energy and raw materials that are used in the entire life cycle of a product. This method lacks a universal application making it insufficient for a cross-segment, region and industry implementation which is needed to create a reporting standard for all companies.

A method that uses a top-down approach is the method composed of financial statements (MC3) as developed by Carballo-Penela and Doménech (2010). It first calculates the organization's carbon footprint and then divides this over products or services offered by the company. Cagiao et al. (2011) state that contrary to the P-LCA method, it allows for a single methodology for organizations and products. The method uses financial statements as input data resulting that no product or process may be omitted. Lastly, the method is comparable across segments, regions and industries since the scope of every analysis is the same. This allows the method to be set as a possible standard method to calculate emissions. However, according to Alvarez et al. (2014) comparing results should be done with caution since many factors influence the results such as geographic location, capacity factor and system boundaries which the model does not take into consideration.

The third type of method is the input-output analysis (IO-LCA) developed by Leontief (1970). The method originally aims to understand interactions between economic sectors, producers and consumers. However, the method is complemented to analyze corporate carbon footprints (Wiedmann (2009)). Over the last decades, there has been a vast increase in research done on analytical models based on this method. Schneider (2009) found, however, that the method is not used by countries or corporations. This is due to the fact that a high level of specialization is required which makes it inaccessible to small and medium-sized enterprises restricting it to academic usage only (Cagiao et al. (2011)). The former three methods show different options to calculate the corporate carbon footprint. However, no method is able to calculate the corporate carbon footprint across regions, sectors and industries without disadvantages making it inapplicable as general calculation method rising the need for a new method.

C Variable overview

In the below table, the overview of the different variables included in the framework to estimate the corporate carbon footprint are displayed.

Table 5: Variable overview with description

Variable	Description
Country of Headquarters	Country were the headquarters is located
GICS Industry code	GICS Sector - 2-digit codes (11 groups)
GICS Industry Group Code	GICS Group - 4-digit codes (24 groups)
Scope 1 emission	Direct GHG emissions via production process
Scope 2 emission	Indirect GHG emissions via energy consumption
Scope 3 emission	Indirect GHG emissions in upstream and downstream processes
Scope 1+2 emission	Aggregated Scope 1 and Scope 2 emissions
Estimation method	Estimation method when no information is disclosed
CO2 regulation	Current status on CO2 laws country headquarters
Revenue total	Annual revenue in the reporting year
EBITDA	Earnings Before Interest, Taxes, Depreciation and Amortization
Capital Expenditures	Capital expenditures per firm per reporting year
Plant, Property and Equipment net	Net property, plant, equipment at reporting year-end
Accumulated Depreciation	Accumulated Depreciation at reporting year-end
Plant, Property and Equipment age	Gross PPE divided by depreciation expense per firm at reporting year-end
Capital intensity	Gross PPE per firm divided by revenue at reporting year-end
Intangibles	Intangibles assets per firm per reporting year
Cost of revenue	Costs of revenue per reporting year
Gross margin	Gross margin per reporting year (%)
Assets total	Total Assets of the company at reporting year-end
Long term debt total	Long term debt at reporting year-end
Leverage	Long-term debt divided by total assets at reporting year-end
FTE	Number of employees at reporting year-end
Fuel Intensity	Carbon intensity of the national fuel combustion in ($kgCO^2$-e/kWh)

Note. An overview of the variables used in this study is listed above together with a brief explanation of the variables.

D GICS sector- and industry-codes

In this section, the GICS industries are displayed representing 20 industries divided over 10 sectors

Table 6: GICS sector and GICS industry codes

Sector		**Industry Group**	
10	Energy	1010	Energy
15	Materials	1510	Materials
		2510	Automobiles & Components
		2520	Consumer Durables & Apparel
25	Consumer Discretionary	2530	Consumer Services
		2540	Media
		2550	Retailing
		3010	Food & Staples Retailing
30	Consumer Staples	3020	Food, Beverage & Tobacco
		3030	Household & Personal Products
35	Health Care	3510	Health Care Equipment & Services
		3520	Pharmaceuticals, Biotechnology & Life Sciences
40	Financials	4010	Banks
		4020	Diversified Financials
		4510	Software & Services
45	Information Technology	4520	Technology Hardware & Equipment
		4530	Semiconductors & Semiconductor Equipment
50	Telecommunication Services	5010	Telecommunication Services
55	Utilities	5510	Utilities
60	Real Estate	6010	Real Estate

Note. In the table, the GICS sector and GICS industry codes are displayed.

E Descriptive statistics variables after pre-processing

The following two tables display descriptive statistics on both data sets. Note that the continuous variables (except the fuel intensity variable) have been log-transformed to remove the skewness. The first table shows continuous variables where the second one displays the categorical variables.

Table 7: Descriptive statistics after pre-processing

Variable	**# Obs**	**Mean**	**Median**	**Std Error**	**Kurtosis**	**Skewness**
Carbon footprint data						
LogCE1	15793	-2.470	-2.517	3.057	-0.225	-0.084
LogCE2	15793	-2.276	-2.105	2.191	0.723	-0.583
LogCE3	8979	-2.371	-2.910	3.359	-0.236	0.347
LogCE12	15793	-1.233	-1.248	2.429	-0.016	-0.145
LogRevenue	15793	8.532	8.546	1.488	0.075	-0.189
LogEBITDA	15793	6.763	6.739	1.421	0.232	-0.022
LogEBIT	15793	6.352	6.330	1.462	0.399	-0.103
LogCapEx	15793	5.525	5.610	1.813	1.370	-0.565
LogPPE_Age	15793	2.435	2.590	0.891	20.027	-2.709
LogPPE_Net	15793	7.078	7.294	2.141	1.645	-0.924
LogIntang	15793	0.970	0.000	2.618	5.648	0.509
LogAsset_tot	15793	9.077	9.019	1.479	-0.003	0.203
LogLTDebt_tot	15793	7.093	7.327	2.050	3.313	-1.274
LogFTE	15793	9.377	9.473	1.618	0.886	-0.581
GrossMar (%)	15793	0.485	0.426	0.285	-0.971	0.428
Leverage (%)	15793	0.202	0.186	0.152	1.892	0.941
Capi_Inten (%)	15598	1.143	0.554	1.674	19.428	3.682
FuelIntensity ($kgCO^2$-e/kWh)	15793	404.092	414.115	219.163	0.085	0.311
Equity value data						
LogMarCap	14582	22.681	22.667	1.333	-0.056	0.075
LogBookVal	14582	21.896	21.890	1.356	0.193	-0.113
LogIncomeNet	14582	20.084	20.070	1.485	0.698	-0.181
LogLiabTot	14582	22.328	22.310	1.643	0.055	0.100
LogOperatingIncome	14582	20.209	20.182	1.415	0.165	-0.044

Note. In the table, the descriptive statistics of 2834 companies with 15.793 observations from 2007-2018 are shown of the data after pre-processing for the corporate carbon footprint model. The data used for the equity model contains 14.582 observations with 2629 companies. The prefix 'log' means that the skewed variable is log transposed. Please read Section 3 for information on missing values and pre-processing steps.

Table 8: Descriptive statistics dummy after pre-processing

Variable	# Obs	Mean	Median	Std Error	Kurtosis	Skewness
GICS Sectors						
Energy	970	0.062	N.A.	N.A.	N.A.	N.A.
Materials	2189	0.140	N.A.	N.A.	N.A.	N.A.
Industrials	3136	0.201	N.A.	N.A.	N.A.	N.A.
Consumer Discretionary	1782	0.114	N.A.	N.A.	N.A.	N.A.
Consumer Staples	1295	0.083	N.A.	N.A.	N.A.	N.A.
Health Care	823	0.053	N.A.	N.A.	N.A.	N.A.
Financials	1199	0.077	N.A.	N.A.	N.A.	N.A.
Information Technology	1450	0.093	N.A.	N.A.	N.A.	N.A.
Communication Services	975	0.063	N.A.	N.A.	N.A.	N.A.
Utilities	930	0.060	N.A.	N.A.	N.A.	N.A.
Real Estate	851	0.055	N.A.	N.A.	N.A.	N.A.
CO2 regulation						
Nationally Implemented	3421	0.219	N.A.	N.A.	N.A.	N.A.
Regionally Implemented	1381	0.089	N.A.	N.A.	N.A.	N.A.
Sub-nationally Implemented	988	0.063	N.A.	N.A.	N.A.	N.A.
No law on CO2	6034	0.387	N.A.	N.A.	N.A.	N.A.
Other	3776	0.242	N.A.	N.A.	N.A.	N.A.
Income Group						
High	9613	0.616	N.A.	N.A.	N.A.	N.A.
Upper Middle	1812	0.116	N.A.	N.A.	N.A.	N.A.
Lower Middle	399	0.026	N.A.	N.A.	N.A.	N.A.

Note. In the table, the descriptive statistics of 2834 companies with 15.793 observations from 2007-2018 are shown of the corporate carbon footprint data after pre-processing. All data on GICS Sectors, CO2 Regulation and Income Group are one-hot encoded. Please read Chapter Data for information on missing values and pre-processing steps.

F Predictor variable correlation

This section shows the correlation matrix for the log-transformed continuous variables. Multiple predictors are highly correlated, where predictors belonging to the group scale of operations are correlated most.

Table 9: Correlation predictors corporate carbon footprint model

	Rev_tot	EBITDA	CapEx	PPE_Age	PPE_Net	Intang	Asset_tot	LTDebt_tot	FTE	GrossMar	Leverage	Capi_Inten	FuelIntensity
Rev_tot	1	0.807	0.706	0.037	0.648	0.159	0.789	0.543	0.743	-0.136	-0.026	-0.179	-0.049
EBITDA	0.807	1	0.762	0.018	0.659	0.132	0.854	0.646	0.596	0.089	0.102	0.027	-0.082
CapEx	0.706	0.762	1	0.192	0.836	0.115	0.655	0.551	0.582	-0.137	0.160	0.236	-0.044
PPE_Age	0.037	0.018	0.192	1	0.411	-0.029	-0.007	0.053	0.031	-0.286	0.076	0.403	0.058
PPE_Net	0.648	0.659	0.836	0.411	1	0.104	0.558	0.487	0.542	-0.250	0.169	0.335	0.010
Intang	0.159	0.132	0.115	-0.029	0.104	1	0.171	0.099	0.112	-0.002	-0.043	-0.036	0.020
Asset_tot	0.789	0.854	0.655	-0.007	0.558	0.171	1	0.714	0.526	0.129	0.067	0.016	-0.081
LTDebt_tot	0.543	0.646	0.551	0.053	0.487	0.099	0.714	1	0.348	0.059	0.499	0.135	-0.109
FTE	0.743	0.596	0.582	0.031	0.542	0.112	0.526	0.348	1	-0.230	-0.060	-0.237	0.019
GrossMar	-0.136	0.089	-0.137	-0.286	-0.250	-0.002	0.129	0.059	-0.230	1	0.008	0.073	-0.080
Leverage	-0.026	0.102	0.160	0.076	0.169	-0.043	0.067	0.499	-0.060	0.008	1	0.248	-0.078
Capi_Inten	-0.179	0.027	0.236	0.403	0.335	-0.036	0.016	0.135	-0.237	0.073	0.248	1	-0.024
FuelIntensity	-0.049	-0.082	-0.044	0.058	0.010	0.020	-0.081	-0.109	0.019	-0.080	-0.078	-0.024	1

Note. In the table, the correlation between the log-transformed continuous predictors used in the corporate carbon footprint model is displayed. Note the high correlation between the predictors belonging to the group scale of operations. The correlation between the different scopes of emissions can be found in the next table.

Table 10: Correlation predictors equity value model

	Scope 1	**Scope 2**	**Scope 1+2**	**Total assets**	**Bookvalue**	**Net income**	**Total liabilities**	**Operating income**
Scope 1	1	0.661	0.919	0.383	0.424	0.370	0.366	0.415
Scope 2	0.661	1	0.832	0.451	0.494	0.456	0.426	0.486
Scope 1+2	0.919	0.832	1	0.452	0.499	0.435	0.429	0.479
Total assets	0.383	0.451	0.452	1	0.904	0.780	0.975	0.822
Bookvalue	0.424	0.494	0.499	0.904	1	0.790	0.811	0.807
Net income	0.370	0.456	0.435	0.780	0.790	1	0.727	0.952
Total liabilities	0.366	0.426	0.429	0.975	0.811	0.727	1	0.781
Operating income	0.415	0.486	0.479	0.822	0.807	0.952	0.781	1

Note. In the table, the correlation between the log-transformed continuous predictors used in the equity value model is displayed. Note the high correlation between the predictors belonging to the group scale of operations.

G Hyperparameter tuning

This appendix shows the optimal hyperparameters after tuning. The ranges are based on findings in Nguyen et al. (2021) and other tuning processes where the same algorithms are used on different data sets. This study uses a sequential gridsearch to find the optimal result since a complete gridsearch over all possible parameters is too computational expensive.

Table 11: Hyperparameters after tuning corporate carbon footprint model

Model	Tuning Parameter	Stand-ardized Variable	Scope 1	Scope 2	Scope 1+2
Linear methods					
OLS	Fit Intercept FI: choice (True, False)	Yes	FI = True	FI = True	FI = True
Lasso	Fit Intercept FI: choice (True, False) Regularization alpha α: uniform (0,1)	Yes	FI = True α = 0.001	FI = True α = 0.001	FI = True α = 0.001
Ridge	Fit Intercept FI: choice (True, False) Regularization alpha α: uniform (0,1)	Yes	FI = True α = 0.01	FI = True α = 10	FI = True α = 0.001
Elastic Net	Fit Intercept FI: choice (True, False) Regularization alpha α: uniform (0,1) L1 ratio L1: unfirm (0,1)	Yes	FI = True α = 0.0001 L1 = 0.49	FI = True α = 0.001 L1 = 0.49	FI = True α = 0.0001 L1 = 0.49
Ensemble methods					
Random Forest	The # of trees in the forest T: range (500,3000, step=1) Max features F: choice (auto, square root, log2) Max depth of the tree D: range (2,3,4,5,6,7,8,9,10,20,30, 40, 50, ..., 100) Min samples to split an internal node S: range (2,5,10,20,30,50,100) Min samples to be at a leaf node L: range (1,2,4, 10,20,30,50)	No	T = 2500 F = sqrt D = 80 S = 2 L = 1	T = 2000 F = sqrt D = 100 S = 2 L = 1	T = 2500 F = sqrt D = 30 S = 2 L = 1
LightGBM	Maximum # tree leaves in tree X: range(100, 200, ..., 1500) Max depth of the tree D: range (5, 6, ..., 12) Min samples to be at a leaf node LE: range (2, 3, ..., 8) Learning Rate L: range (0.00001, 0.0001, 0.005, 0.001, 0.05, 0.01,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0,9,1) Subsample ratio SS: range (0.001, 0.005, 0.01, 0.1, 0.2, 0.3) Column sample ratio C: range (0.5, 1, step =0.1)	No	X = 300 D = 8 LE = 7 L = 0.01 SS = 0.001 C = 0.5	X =300 D = 7 LE = 3 L = 0.01 SS = 0.001 C = 0.5	X =300 D = 8 LE = 6 L = 0.01 SS = 0.001 C = 0.5
XGBoost	Max depth of the tree D: range (9,10,11, 12) The minimum child weight CW: range (5,6,7,8) Learning Rate L: range (0.00001, 0.0001, 0.005, 0.001, 0.05, 0.01, 0.1,0.2, ..., 1) Subsample ratio SS: range (0.5, 1, step =0.1) Column sample ratio C: range (0.1, 1, step =0.1) The # of boosting rounds B: (1000)	No	D = 4 CW = 7 L = 0.1 SS = 0.9 C = 0.9 B = 1000	D = 4 CW = 7 L = 0.1 SS = 0.9 C = 0.9 B = 1000	D = 4 CW = 9 L = 0.1 SS = 1 C = 0.6 B = 1000
Non-Linear Algortihms					
Neural Network	Hidden layer sizes: H1 choice (10,20, ..., 100) Hidden layer sizes: H2 choice (0,10, ..., 100) Activation function A: choice (identity, relu, sigmoid, tanh) Learning Rate L: choice (constant, inverse scaling, adaptive) Initiatal Learning Rate LI: uniform (0,1) Regularization alpha α : uniform (0,1) Max Iterations M range (1000,2500, step= 500)	Yes	H1 = 70 H2 = 0 A = Tanh L = Constant L1 = 0.00009 α = 0.10469 M = 2000	H1 = 40 H2 = 0 A = logistic L = adaptive L1 = 0.0002 α = 0.01953 M = 1000	H1 = 90 H2 = 90 A = tanh L = invscaling L1 = 0.00005 α = 0.03544 M = 2000

Note. In this table, the range of the grids and the final optimal hyperparameters are displayed. The name of the hyperparameters come from the respective packages they come from (Scikit Learn, XGBoost and LightGBM). The ranges are based on findings in Nguyen et al. (2021) and other tuning processes where the same algorithms are used on different data sets. This study uses a sequential gridsearch to find the optimal result. The final parameters reflect the result with highest prediction performance (lowest MAE) using 5-fold-group-cross-validation.

Table 12: Hyperparameters after tuning equity value model

Model	Tuning Parameter	Standardized Variable	Scope 1	Scope 2	Scope 1+2
Linear methods					
OLS	Fit Intercept FI: choice (True, False)	Yes	FI = True	FI = True	FI = True
Lasso	Fit Intercept FI: choice (True, False) Regularization alpha α: uniform (0,1)	Yes	FI = True α = 0.001	FI = True α = 0.0001	FI = True α = 0.0001
Ridge	Fit Intercept FI: choice (True, False) Regularization alpha α: uniform (0,1)	Yes	FI = True α = 10	FI = True α = 10	FI = True α = 10
Elastic Net	Fit Intercept FI: choice (True, False) Regularization alpha α: uniform (0,1) L1 ratio L1: unfirm (0,1)	Yes	FI = True α = 0.0001 L1 = 0.49	FI = True α = 0.001 L1 = 0	FI = True α = 0.0001 L1 = 0.32
Ensemble methods					
Random Forest	The # of trees in the forest T: range (500,3000, step=1) Max features F: choice (auto, square root, log2) Max depth of the tree D: range (2,3,4,5,6,7,8,9,10,20,30, 40, 50, ..., 100) Min samples to split an internal node S: range (2,5,10,20,30,50,100) Min samples to be at a leaf node L: range (1,2,4, 10,20,30,50)	No	T = 2500 F = sqrt D = 30 S = 2 L = 1	T = 1000 F = sqrt D = 60 S = 2 L = 1	T = 3000 F = sqrt D = 40 S = 2 L = 1
LightGBM	Maximum # tree leaves in tree X: range(100, 200, . . . , 1500) Max depth of the tree D: range (5, 6, . . . , 12) Min samples to be at a leaf node LE: range (2, 3, . . . , 8) Learning Rate L: range (0.00001, 0.0001, 0.005, 0.001, 0.05, 0.01,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0,9,1) Subsample ratio SS: range (0.001, 0.005, 0.01, 0.1, 0.2, 0.3) Column sample ratio C: range (0.5, 1, step =0.1)	No	X = 300 D = 5 LE = 5 L = 0.01 SS = 0.001 C = 0.9	X = 300 D = 6 LE = 7 L = 0.01 SS = 0.001 C = 0.8	X = 300 D = 6 LE = 7 L = 0.01 SS = 0.001 C = 0.8
XGBoost	Max depth of the tree D: range (9,10,11, 12) The minimum child weight CW: range (5,6,7,8) Learning Rate L: range (0.00001, 0.0001, 0.005, 0.001, 0.05, 0.01, 0.1,0.2, ..., 1) Subsample ratio SS: range (0.5, 1, step =0.1) Column sample ratio C: range (0.1, 1, step =0.1) The # of boosting rounds B: (1000)	No	D = 4 CW = 5 L = 0.1 SS = 0.9 C = 0.7 B = 1000	D = 4 CW = 5 L = 0.1 SS = 0.9 C = 0.7 B = 1000	D = 4 CW = 6 L = 0.05 SS = 1 C = 1 B = 1000
Non-Linear Algortihms					
Neural Network	Hidden layer sizes: H1 choice (10,20, ..., 100) Hidden layer sizes: H2 choice (0,10, ..., 100) Activation function A: choice (identity, relu, sigmoid, tanh) Learning Rate L: choice (constant, inverse scaling, adaptive) Initiatal Learning Rate LI: uniform (0,1) Regularization alpha α : uniform (0,1) Max Iterations M range (1000,2500, step= 500)	Yes	H1 = 90 H2 = 90 A = Logistic L = Invscaling L1 = 0.001 α = 0.051 M = 2000	H1 = 50 H2 = 0 A = Logistic L = Adaptive L1 = 0.001 α = 0.039 M = 2500	H1 = 60 H2 = 0 A = Logistic L = Constant L1 = 0.006 α = 0.021 M = 1000

Note. In this table, the range of the grids and the final optimal hyperparameters are displayed. The name of the hyperparameters come from the respective packages they come from (Scikit Learn, XGBoost and LightGBM). The ranges are based on findings in Nguyen et al. (2021) and other tuning processes where the same algorithms are used on different data sets. This study uses a sequential gridsearch to find the optimal result since a complete gridsearch over all possible parameters is too computational expensive. The final parameters reflect the result with highest prediction performance (lowest MAE) using 5-fold-group-cross-validation.

H Model Confidence Set

Table 13: P-values Model Confidence Set corporate carbon footprint models

P-values	Scope 1	Scope 2	Scope 1+2
OLS	0.000	0.000	0.000
Ridge	0.000	0.000	0.000
Lasso	0.000	0.000	0.000
Elastic Net	0.000	0.000	0.000
Random Forest	0.000	0.007	0.000
LigthGBM	**1.000**	**1.000**	**1.000**
XGBoost	0.000	0.000	0.000
Neural Network	0.020	0.007	0.000

Note. In the table, the p-values of the Model Confidence Set are displayed. A p-value >0.05 results in inclusion of the respective method in the significantly superior set. Here, a p-value printed in bold means inclusion in the superior set. The methods are estimated using the hyperparameters that are found after tuning.

Table 14: P-values Model Confidence Set equity value models

P-values	Scope 1	Scope 2	Scope 1+2
OLS	0.000	0.000	0.000
Ridge	0.000	0.000	0.000
Lasso	0.000	0.000	0.000
Elastic Net	0.000	0.000	0.000
Random Forest	0.000	0.000	0.000
LigthGBM	0.000	**0.333**	0.000
XGBoost	0.000	0.000	0.000
Neural Network	**1.000**	**1.000**	**1.000**

Note. In the table, the p-values of the Model Confidence Set are displayed. A p-value >0.05 results in inclusion of the respective method in the significantly superior set. Here, a p-value printed in bold means inclusion in the superior set. The methods are estimated using the hyperparameters that are found after tuning.

I Code of the thesis

Please find the code used in this thesis attached in a ZIP file. This folder contains twelve separate scripts with the following content:

Filename	**Description**
Data pre-processen copy	This code is used to pre-process the data.
Linear Models CE	This code performs OLS, ridge, Lasso and Elastic Net for the corporate carbon footprint model.
Linear Models MarCap	This code performs OLS, ridge, Lasso and Elastic Net for the equity value model.
Ensemble Methods CE	This code performs the Random Forest, Extreme Gradient Boosting and LightGBM algorithm for the corporate carbon footprint model.
Ensemble Methods MarCap	This code performs the Random Forest, Extreme Gradient Boosting and LightGBM algorithm for the equity value model.
Neural Networks CE	This code performs the Neural Network for the corporate carbon footprint model.
Neural Networks MarCap	This code performs the Neural Network for the equity value model.
SHAP values CE	This code computes the SHAP values for the best performing method for the corporate carbon footprint model.
SHAP Values MarCap	This code computes the SHAP values for the best performing method for the equity value model.
Model Confidence Set CE	This code computes the MCS for the corporate carbon footprint model.
Model Confidence Set MarCap	This code computes the MCS for the equity value model.
Jacobian Matrix NN MarCap	This code performs a sensitivity analysis for the best performing neural networks for the equity value model.

Note. Overview of the attached files.