Firms' profitability and ESG score: a machine learning approach

Rita L. D'Ecclesia, Valeria D'Amato, Susanna Levantesi

June 3, 2024

SAPIENZA
UNIVERSITÀ DI ROMA

Agenda

- Motivation
- Aim
- Literature review
- Methodology
- Data
- Results
- Concluding Remarks

Motivation

Motivation

- Regulators's requirements and government's policies in terms of Sustainability are affecting the investment environment and the challenges and opportunities faced by companies.
- The ESG rating and its use in the investment choices has become a mantra.
- Stranded assets, particularly from climate-related physical and transition risks, has spurred work by financial supervisors and central banks.
- Recent regulatory directives require mandatory disclosure of sustainable activity in some countries, only on a voluntary basis in others.

Motivation

- Financial institutions and their supervisors are still at an early stage in developing and deploying suitable datasets, models, and tools.
- Better data and analysis to properly measure and manage exposures to environment-related risks are required.
- Existing literature pointed out that corporate social responsibility (CSR) has a potential impact on the performance of the firm.
- There is only limited evidence of the relationship between non-financial indicators (e.g. ESG score) and the firm's profitability

Sustainable and Responsible Investments

- The European Sustainable Investment Forum (Eurosif) defines "Sustainable and responsible investment (SRI)" as a long-term oriented investment approach which integrates ESG factors in the research, analysis and selection process of securities within an investment portfolio.
- S&R finance has become a major aim for asset managers who are regularly dealing with the measurement and management of ESG risks.
- 3 primary drivers of increased ESG investment:
  - sustainability challenges,
  - shifts in investor preferences,
  - improvement in data and analytics.

Motivation

- In 2022 total U.S.-domiciled sustainably invested assets under management (institutional + retail), dropped to $8.14 trillion, representing 12% of the $66 trillion in total U.S. AUM (Forum for Sustainable and Responsible Investment’s 2022 trends report), (while in 2020 they represented 33%)
- Active strategies represent the majority of ESG-related AUM, at 60% in the U.S. and 82% in Europe.
- In this context asset managers look for some assessment of sustainability for guidance and benchmarking
- ESG score aimed to provide disclosure on the E,S and G (corporate social responsibilities) metrics.
- CSR/ESG ratings are becoming quite popular even if highly questioned in terms of reliability

Aim

Aim of the paper

- To assess how structural data- balance sheet items may be related to ESG scores assigned to regularly traded stocks.
- We focus on ESG investments by considering the ESG scores for explaining the firms' profitability, not trivial the virtuous circle between ESG investments and the firms�much success.
- Using a ML approach (Random Forest algorithm), we investigate how a firm’s profitability may be affected by ESG scores for the companies which constitute the STOXX 600 Index.
- We study the relationship between ESG score and EBIT using machine learning interpretability tool-boxes

Literature Review

Literature Review 1.1

Financial performance of "sustainable companies":

- Higher future returns for better ESG firms
  - Mahjoub and Khamoussi, 2012; Mahler et al.,2009; Trucost and Mercer, 2010; Nakao et al., 2007; Weber et al., 2010; Derwall et al., 2005; Van de Velde et al., 2005
- Lower future returns for better ESG firms
  - Makni et al., 2009; Renneboog et al.2008; Simpson and Kohe2002; Angel et al., 1997;
- No meaningful differences
  - Belghitar et al. (2014), Hamilton et al. (1993), Statman (2000), Bauer et al. (2005), Bello (2005), Kreander et al. (2005), Utz et al. (2014)).

Literature Review 1.2

<u>Performance of Sustainable Companies</u>

- Companies with robust ESG practices display a lower cost of capital, lower volatility and few instances of bribery, corruption and fraud, the opposite happening for companies with poor ESG practices (Lins et al., 2017; Chava, 2011; Lansilahti, 2012; Bhagat et al., 2008; Cremers et al., 2005; Deutsche Bank, 2012).
- Contradicting results Arribas et al. (2019a): the unexpected performance of sustainable and conventional mutual funds are mainly due to the way you measure sustainability
- (Lin et al., 2019) bidirectional linkages between CSR and corporate financial performance (CFP) Panel Vector Autoregression in GMM context.
- Drempetic et al. (2020) the **availability** of a company’s ESG data on the company’s sustainability and fin. performance are positively correlated

Literature Review 1.3

Profitability and ESG

- the relationship between structural data, i.e. balance sheet and income statement information, and CSR performance as in Drempetic et al. (2020), Garcia et al. (2020), and Lin et al. (2019). They use vague and heterogeneous data-sets so the results are not robust and have to use the Rough data set theory.
- The theory of slack resources is used: profitability is expected to have a positive impact on the ESG score: companies with the greatest resources are precisely those who can afford the necessary investments to improve the ESG score (Drempetic et al., 2020).

Literature Review 1.4

- Green bonds, largely issued by corporations, play a key role even if they have provided the same risk-return profiles of its conventionally counterpart (Hachenberg and Schiereck, 2018). Pricing of green bonds does not reflect the quality of the bond (Zerbib, 2019).

Literature Review 2.1

- Poor consistence of ESG rating:Jewell et al. (1998): credit ratings from Moody’s and S&P’s are correlated at 0.99, (Berg et al. (2019) while correlation between ESG ratings on average 0.61).
- Chatterji et al. (2016): ratings from different providers differ dramatically, information received from CRAs is quite noisy.

Literature Review 2.2

- Melas et al., 2018 explain how ESG characteristics have led to financially significant effects:
  - Borrowing from Central Banks creates 3 “transmission channels” within a standard DCF model:
    - i) the cash-flow channel,
    - ii) the idiosyncratic risk channel
    - iii) the valuation channel.
  - i) and ii) are transmitted through corporations’ idiosyncratic risk profiles,
  - iii) is linked to companies’ systematic risk profiles.

  ESG has an effect on valuation and performance…………..

Literature Review 3.1

- Multivariate approach
  - Weber et al. (2008) employs ESG criteria to predict accounting indicators, using logistic regression. The results indicate that the statistical approach is useful to show that ESG performance can explain corporate financial performance with regard to some EBITDA margin, ROA, and ROE. However, it cannot predict TR, because there might be too many other important influences on TR (Cerin et al., 2001) or that the shareholders do not integrate.
  - Weber (2017) investigates the connection between ESG performance and financial performance of Chinese banks used panel regression. the integration of ESG data is useful for financial decision makers (Monk et al., 2019).

Literature Review 3.2

- Artificial Intelligence approach
  - In et al., 2019 : AI may improve the collection and data as well as its analysis
  - ESG data can be text data, categorical data, and quantitative so AI methods are able to recognize patterns without assuming a certain distribution of the data
  - Wang et al. (2012) use the AdaBoost algorithm to forecast equity returns and Wang et al. (2014) show that using different training windows provides better performance. Batres-Estrada (2015) and Takeuchi and Lee (2013) Moritz and Zimmermann (2016) use tree-based models to predict portfolio returns. Gu et al. (2020) forecast individual stock returns
  - Alberg and Lipton (2017) who propose to forecast company fundamentals (e.g., earnings or sales) rather than returns.

Methodology

Machine Learning

A generic regression model for estimating the relationship between a target (or response) variable, Y , and a set of predictors (or features), $X_1, X_2, ..., X_p$

$$Y = f(X_1, X_2, ..., X_p) + \epsilon$$

To estimate f we apply both, individual techniques (Decision Trees) and ensemble methods (bagging, gradient boosting, Random forest) by minimizing the reducible error:

$$E[f(X_1, X_2, ..., X_p) - \hat{f}(X_1, X_2, ..., X_p)]^2$$

Comparing ML to the generalized linear model

Decision Trees

- The decision trees (DT) algorithm splits the predictor space X into *J* distinct and non-overlapping regions, providing the same prediction for all the observations falling into each region
- The target variable is estimated by the average values of the variable belonging to the same region *Rj*.

Bagging and Random Forest

- The bagging was designed to improve machine learning algorithms' stability and accuracy.
- It creates multiple bootstrap samples from the training data and fits a weak learner for each sample. The weak learners are aggregated by averaging their outputs.
- Compared to bagging, RF peculiarity is the way of considering the predictors: at each split, RF selects a random subset of predictors as candidates for the subdivision from the final set of predictors, thus preventing the predominance of strong predictors in the subdivisions of each tree (Breiman, 2001).
- RF estimator:

$$\hat{f}^{RF}(\mathbf{x}) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{tree}(\mathbf{X}|b)$$

where B is the number of the bootstrap sample and *f* is the DT estimator over the *b* sample.

Gradient Boosting

- Proposed by Friedman (2001), GB uses fixed-sized DT as weak predictive models (typically, trees with a small number of splits).
- The prediction is obtained with a sequential approach and not parallelizing the tree build process as in RF.
- GB model updating rule:

$$f_m(\mathbf{X}) = f_{m-1}(\mathbf{X}) + \sum_{j=1}^{J_m} \gamma_{jm} \mathbb{1}_{\{\mathbf{x} \in R_{jm}\}}$$

  where γ_{jm} are computed by solving the optimization model $\gamma_{jm} = \arg\min_{\gamma} \sum_{X_j \in R_{jm}} L\left(y_{1j}, f_{m-1}(X_j) + \gamma\right)$, given a specified loss function $L(\cdot)$.
- GB final estimation: $\hat{f}(\mathbf{X}) = f_M(\mathbf{X})$.

GLM

- GLM generalizes linear regression by relating the linear model to the response variable through a link function $g(\cdot)$.
- Linear predictor: $\eta = g(E(Y)) = X\beta$, where β is the vector of the regression coefficients that need to be estimated.
- We assume: Y distributed as a Gaussian and link function $=$ identity, so that: $\eta = E(Y)$.
- Regression model, including three features' interactions:

$$EBIT \sim Year + Net.Sales + ESG.Score + Sector + PE + ROE + DY + I1 + I2 + I3$$

- $I1 = Sector * ESG.Score$
- $I2 = Net.Sales * ESG.Score$
- $I3 = Sector * Net.Sales$

Random Forest: *Main strenght*

- It is a non-parametric method able to catch tricky relations between inputs and outputs, without involving any a-priori assumption.

- It is possible to handle heterogeneous data and intrinsically implement feature selection, making them robust to not significant or noisy variables.

- They are robust to outliers or missing values and are interpretable.

Random Forest: main issues

- The choice of the N of trees should be done carefully, in order to reach the highest % of explained variance and the lowest mean of squared residuals (MSR).
- Variable importance is determined according to the relative influence of each predictor, by measuring the N.of times a predictor is selected for splitting during the tree building process.
- A weighted impurity measure has been proposed in Breiman (2001) for evaluating the importance of a variable Xm in predicting the target *Y* , for all nodes *t* averaged over all *NT* trees in the forest.
- the Gini importance, obtained by assigning the Gini index to the impurity *i*(*t*) index. (Mean Decrease Gini, *MDG)*

$$MDG(X_m) = \frac{1}{N_T} \sum_{T} \sum_{t \in T : v(s_t) = X_m} p(t) \Delta i(s_t, t)$$

Model's Interpretability

- To understand how a ML model operates we need to explain the various stages to know how it works and which decision rules it takes
- Model-agnostic interpretation methods clear up the predictive power of the ML models
  - Partial Dependence Plots (PDP)
  - Accumulated Local Effect plots (ALE)
  - Individual Conditional Expectation (ICE)
  - Ceteris-paribus (CP) profiles
  - Feature interaction
  - SHAP (SHapley Additive exPlanations)

Model's Interpretability

- **PDP** (Friedman, 2001): shows the marginal effect of one or two features entering into the set of the predicted outcome averaged over the joint values of the other input features.
- **ALE** plots (Apley, 2020): shows how the prediction changes locally when the feature is varied. It addresses the bias arising in PD when the selected feature is highly correlated with other features by averaging over a conditional distribution (instead of over a marginal distribution as in PDP).
- **ICE** (Goldstein et al., 2015): disaggregates the output of PDP by providing a certain number of estimated conditional expectation curves.

Model's Interpretability

- **CP profiles**: extension of PDP and ICE plots. CP assesses the influence of a selected feature by assuming that the values of all the other features remain unchanged. The CP profile shows the dependence of the conditional expectation of the target variable on the values of the selected feature.
- **Feature interaction**: measures how strongly features interact with each other (how much of the variance of the model's estimation of the target variable is explained by the interaction). The measure is between 0 (no interaction) and 1 (= 100% of variance of the estimated target variable due to interactions).
- **SHAP**: an alternative method for unfolding individual predictions originates from the coalitional game theory through the Shapley value. It is assumed that, for one observation, the feature values play a game together, in which they get the prediction as a payoff (the model output). The Shapley value shows how to fairly allocate the payoff among the input features.

DATASET

ESG Scores

- Fitch Ratings launched ESG Relevance Scores (ESG.RS) for 1,534 corporate issuers in January 2019; more than 143,000 ESG.RS for over 10,200 issuers and transactions.
- MSCI introduced: ratings, indexes and analytics.
  - Ratings: 75,000 companies and more than 650,000 equity and fixed income securities. based on the exposure of each company to industry specific ESG risks
- The Bloomberg ESG Data Service analyses 11,500 companies in 83 countries. 800 metrics covering all the aspects of ESG, from emissions to the % of women employees. Disclosures: E; S; G

Dataset

400 Companies listed in the STOXX Europe 600 index:

- Balance sheet information
- ESG scores
- E score
- S scores
- G scores

Period: 2011-2020

Variables

Variable	Description
Target variables Y	
EBIT	Earnings Before Interest and Taxes
Features $X_1, X_2, ..., X_p$	
ESG.Score	Measure of the overall corporate social responsibility
Year	2011-2020
Sector	Categorical variable indicating the company's industry sector
Net.Sales	Sales receipts for products and services, less cash discounts, trade discounts, excise tax, and sales returns and allowances
PE	Price-to-Earnings, computed as the ratio of fiscal period Price Close to Earnings Per Share Excluding Extraordinary Items
ROE	Return On Equity, profitability ratio calculated by dividing a company's net income by total equity of common shares (percentage values)
DY	Dividend Yield, calculated as the Dividends paid per share to the primary common shareholders for the fiscal period divided by the Historical Price Close (percentage values)

Main statistics

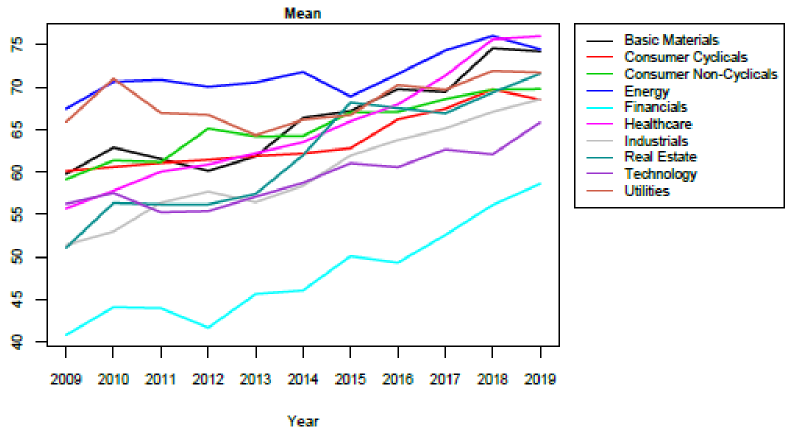


Figure 3: Main statistics of the ESG score distribution by year: average value

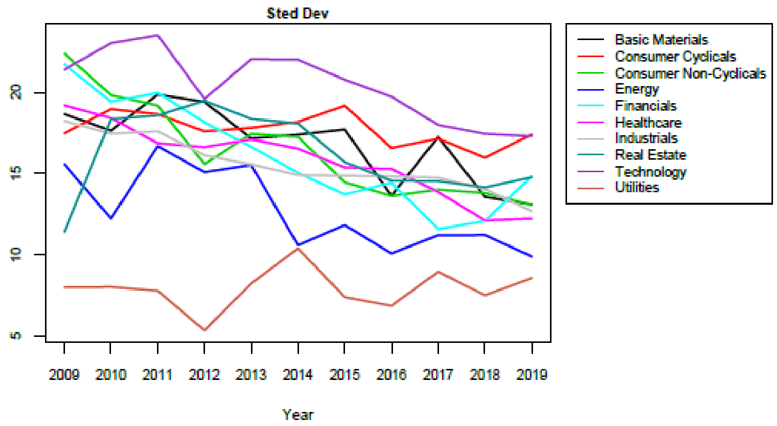


Figure 4: Main statistics of the ESG score distribution by year: standard deviation

Main Features

Figure 1: ESG categories. Source: Thomson Reuters Refinitiv ESG scores

Dataset description

Year	BasMat	ConCyc	ConNCy	Ene	Fin	Hea	Ind	ReaEst	Tec	Uti	All
2011	57.19	59.05	60.40	70.22	41.71	59.96	54.95	57.08	54.21	60.63	57.72
2012	60.87	59.24	62.05	69.13	44.19	60.56	55.97	59.34	57.79	61.83	59.06
2013	61.97	58.71	63.39	72.30	43.80	62.07	56.27	59.60	56.77	60.67	59.52
2014	63.95	58.98	66.37	65.08	43.93	62.99	57.67	64.03	56.67	59.64	60.23
2015	67.95	62.50	66.17	69.20	50.74	63.86	61.66	68.42	58.66	64.06	63.31
2016	67.03	64.33	65.89	72.56	50.39	67.56	63.68	68.52	60.68	64.95	64.68
2017	68.90	66.20	68.87	74.46	54.41	73.31	64.83	67.89	62.77	63.18	66.75
2018	71.45	68.40	70.21	76.03	55.67	73.95	66.86	69.30	63.73	66.98	68.69
2019	73.64	69.34	71.75	69.13	58.23	77.21	70.37	71.15	68.21	73.34	70.69
2020	76.31	72.23	74.21	77.63	63.95	79.91	72.61	71.32	70.07	76.77	73.86

Table 4: Mean values of the ESG score by economic sectors. Years 2011-2020.

ESG variables: *density Function*

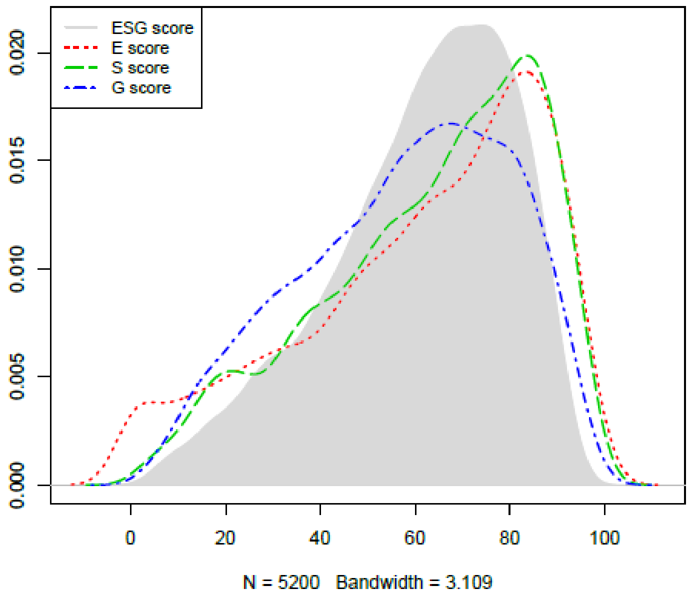


Figure 5: Density functions of ESG variables.

ESG SCORE: Density function by sector

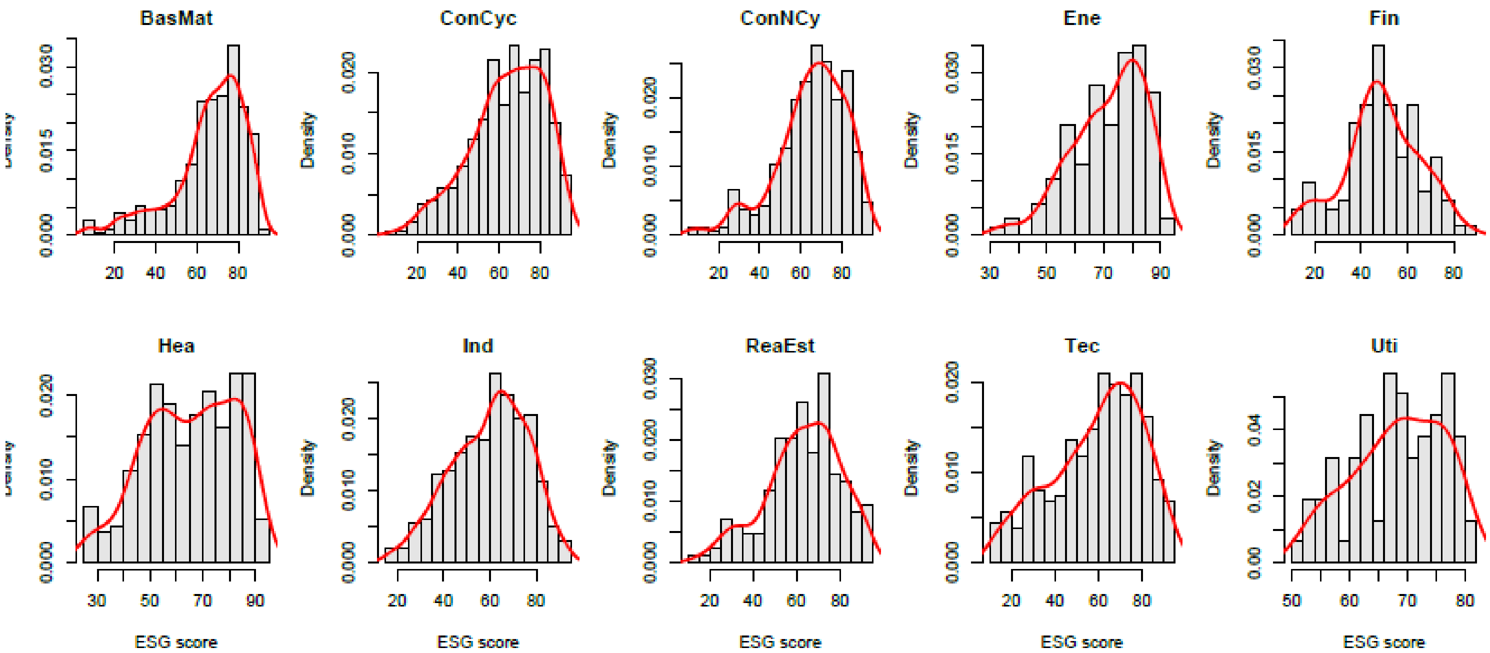


Figure 6: Density functions of ESG.Score by sector.

Dataset description

$EBIT$	$ESG.Score(mean)$
<0	57.02
$0 - 500$	61.33
$501 - 1000$	69.22
$1001 - 2000$	69.99
$2001 - 5000$	71.64
$5001 - 10000$	77.66
> 10000	76.26

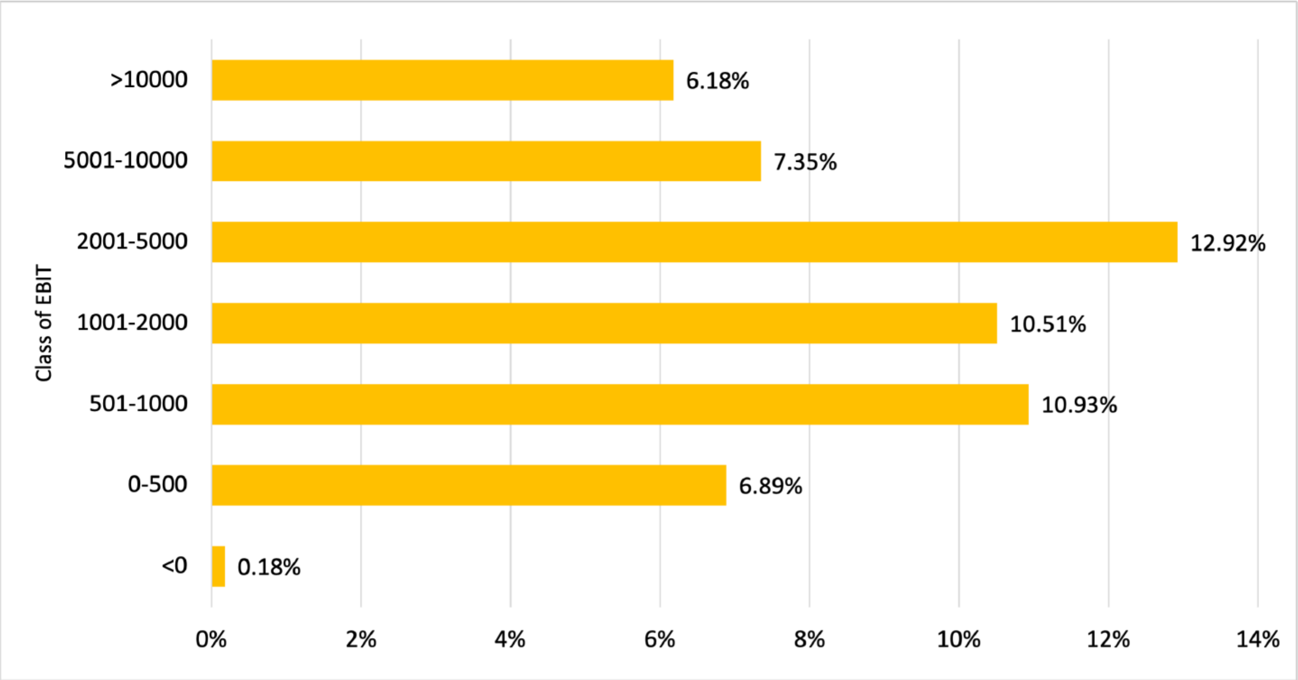


Figure 1: Percentage distribution of EBIT values (in Million Euros). Years 2011-2020.

Results

Model's prediction performance

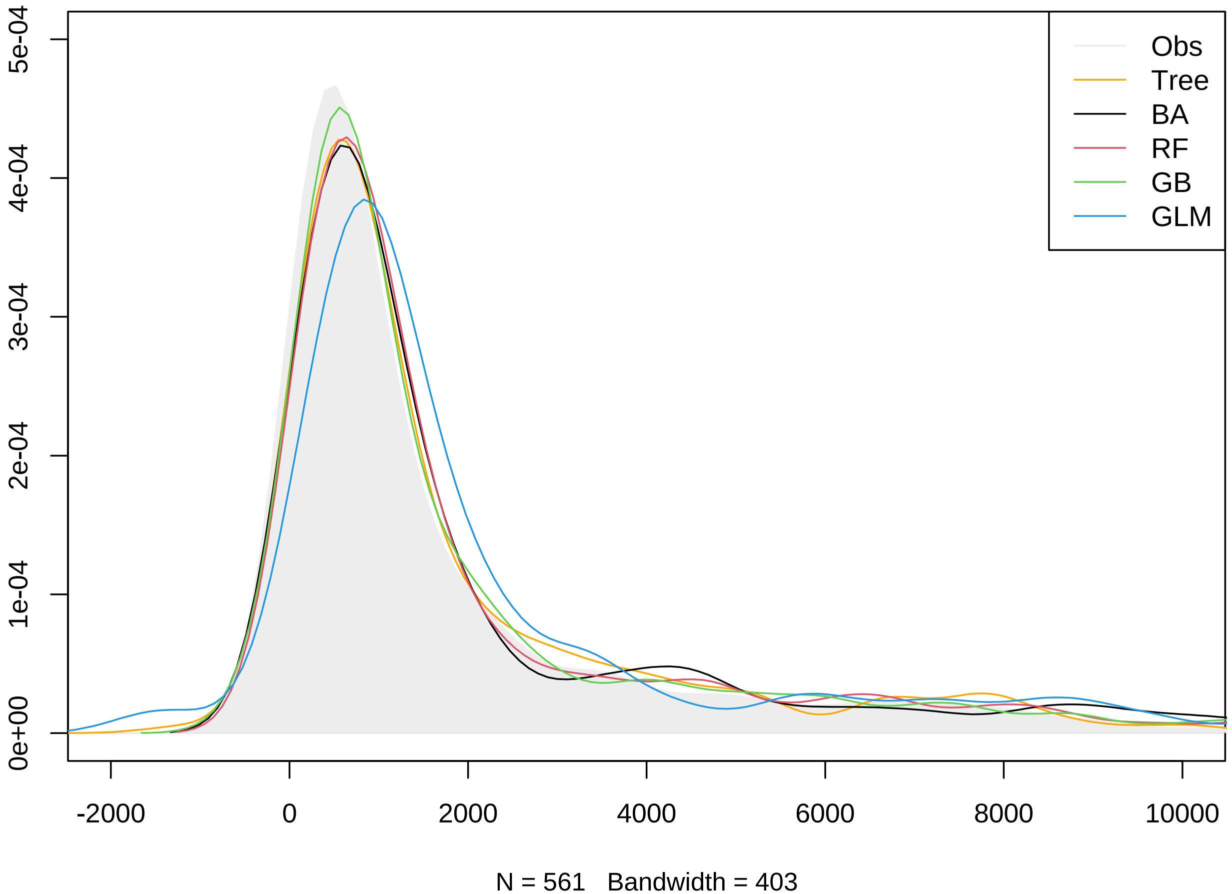
Model	DT	BAG	RF	GB	GLM
R^2	73.18%	87.90%	**88.39%**	88.36%	78.03%

Table 6: R^2 values

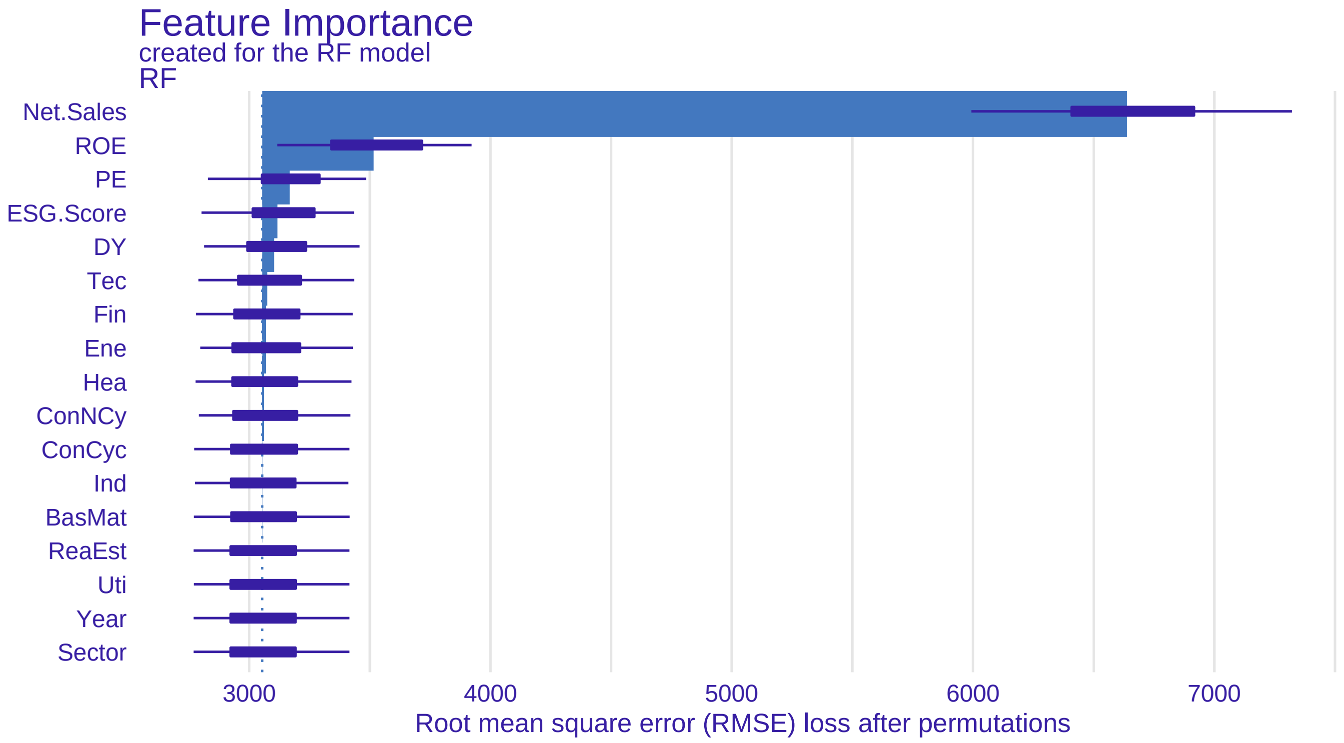
Model	DT	BAG	RF	GB	GLM
RMSE-train	1,808	2,023	1,980	897	2,330
MAE-train	727	831	823	541	1,179
RMSE-test	2,580	2,145	**2,102**	2,104	2,891
MAE-test	1,003	844	**831**	965	1,284

Table 7: RMSE and MAE of EBIT predicted values.

Density functions of observed values and models' estimated values

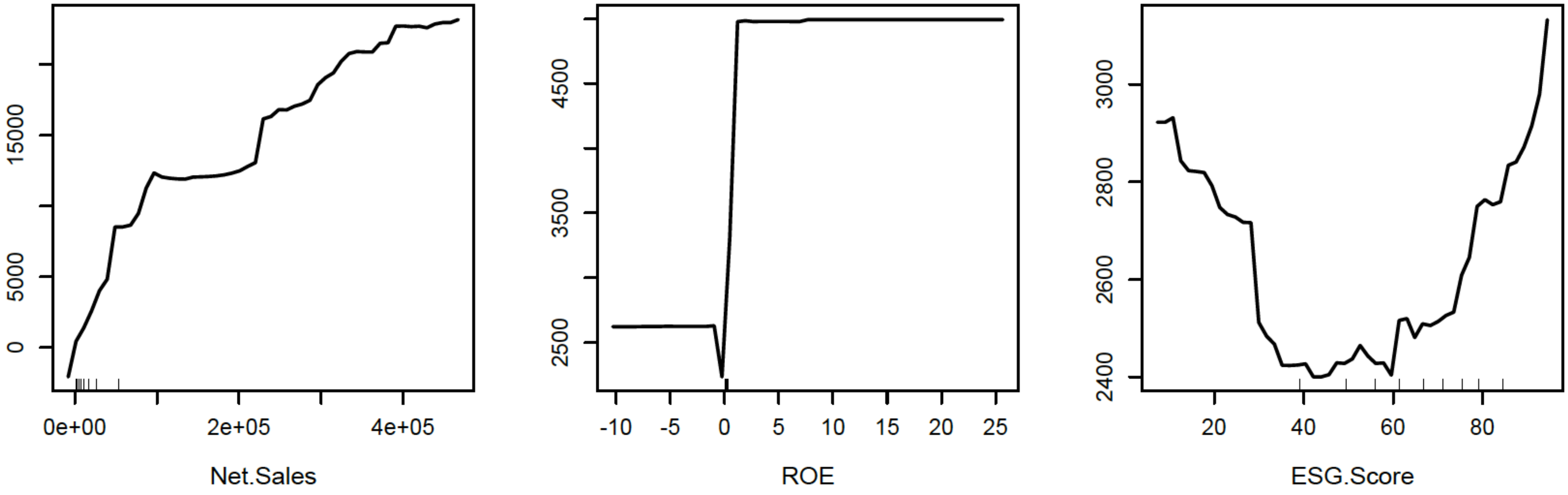


Variable importance



- Most important variable in explaining EBIT: *Net.Sales*, *ROE*, and *ESG.Score*
- We are interested in understanding how the ESG score affects the company's profitability

Partial Dependence plot



- The PDP for the *Net.Sales* shows an increasing trend, as well as for the *ROE* predictor, which reaches a plateau.
- The PDP for the *ESG.Score* could confirm that the insight of lower ESG scores may not necessarily be value-adding, but rather charging the company with other expenses.

Interpretability: ICE plots

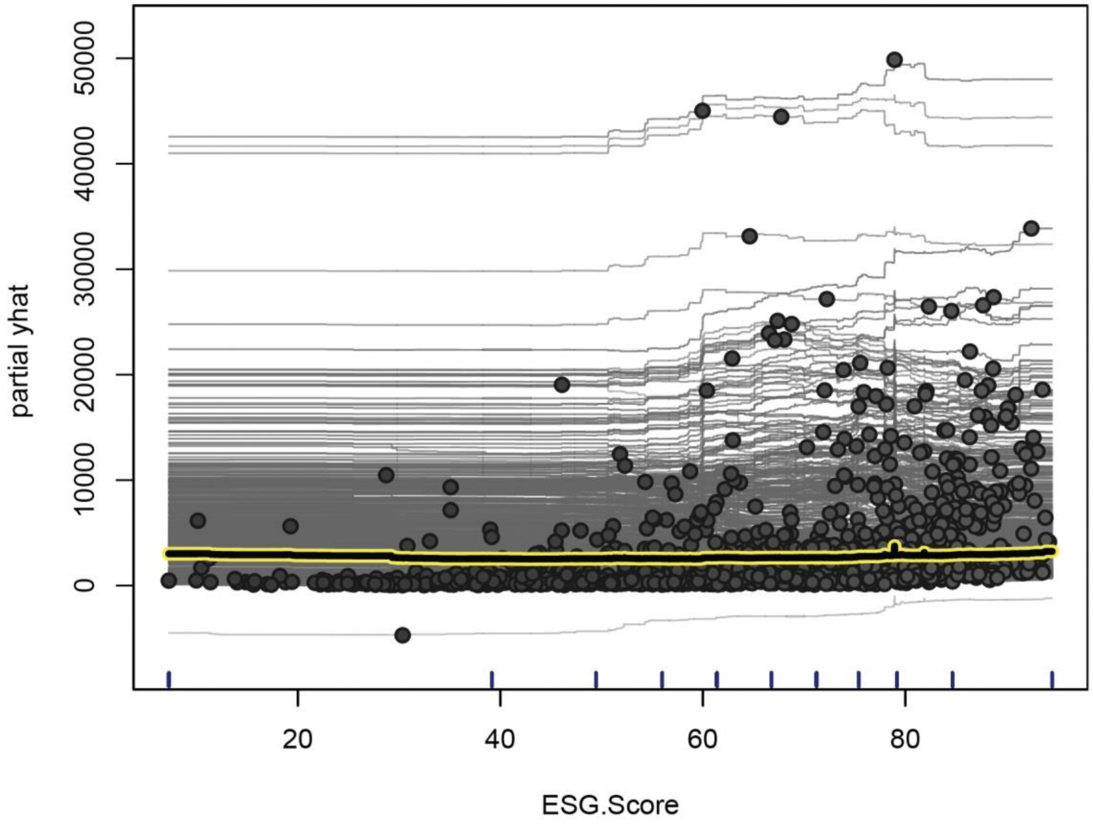


Figure: ICE plot for the ESG score. The yellow line represents the PDP of the ESG score. Each of the grey lines represents the conditional expectation for a single observation

ICE plots highlight the variation in the fitted values across the range of a feature, suggesting where and to what extent heterogeneities might exist (Goldstein et al., 2015)

The predicted EBIT values show a differentiation in the 60-90 range of the ESG score

Interpretability: CP interpretation

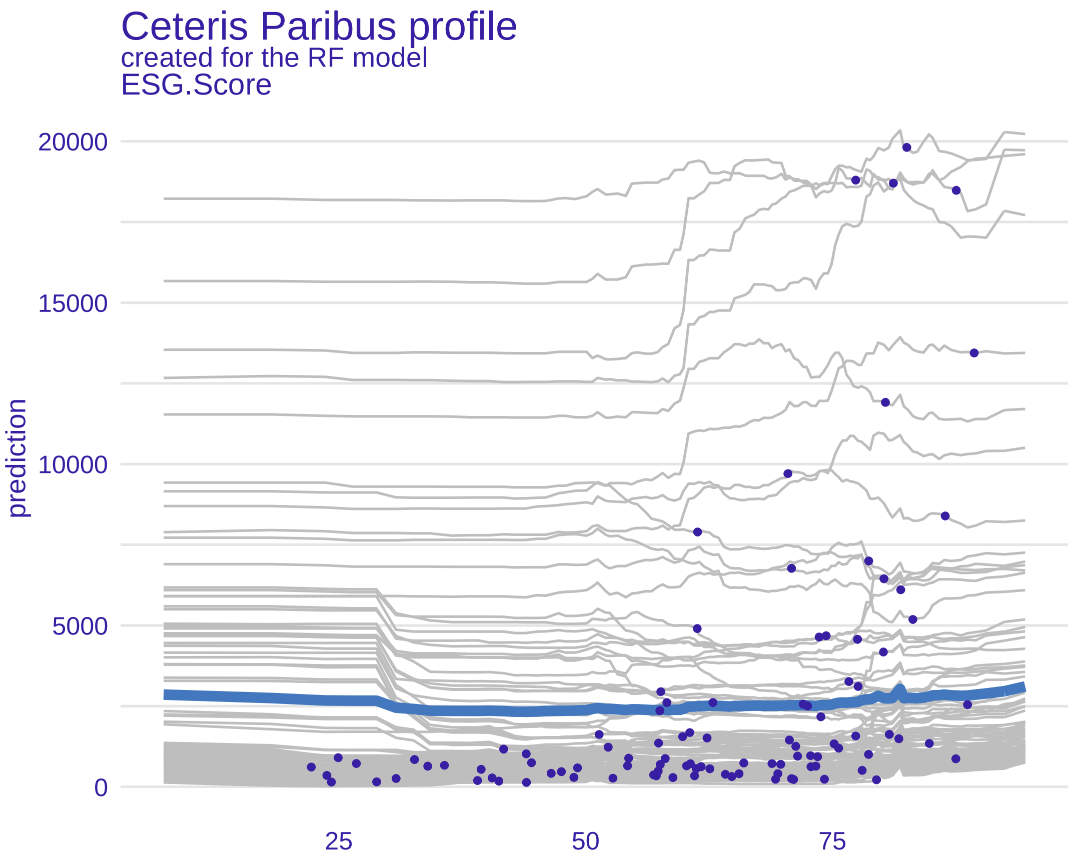
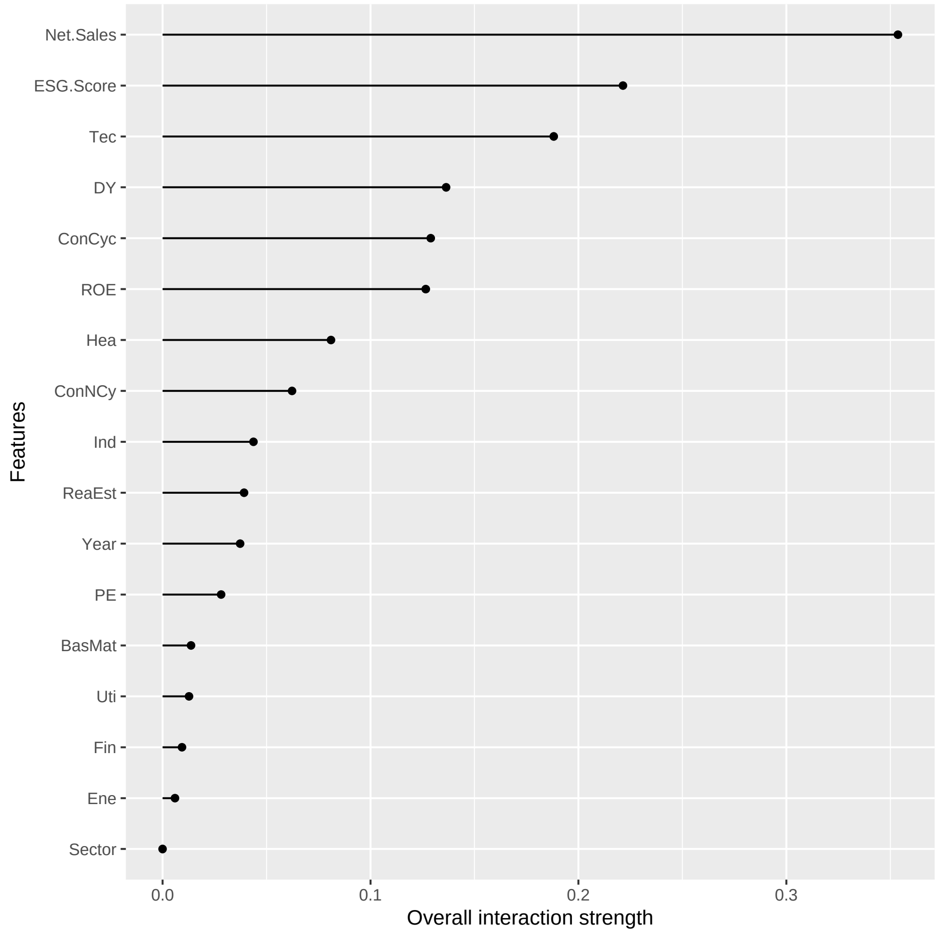


Figure: CP interpretation; features: *ESG.Score*. Grey lines show the CP profiles for 100 randomly selected observations (dark blue dots). The blue line shows the mean of the CP profiles, which offers an estimate of the PD profile.

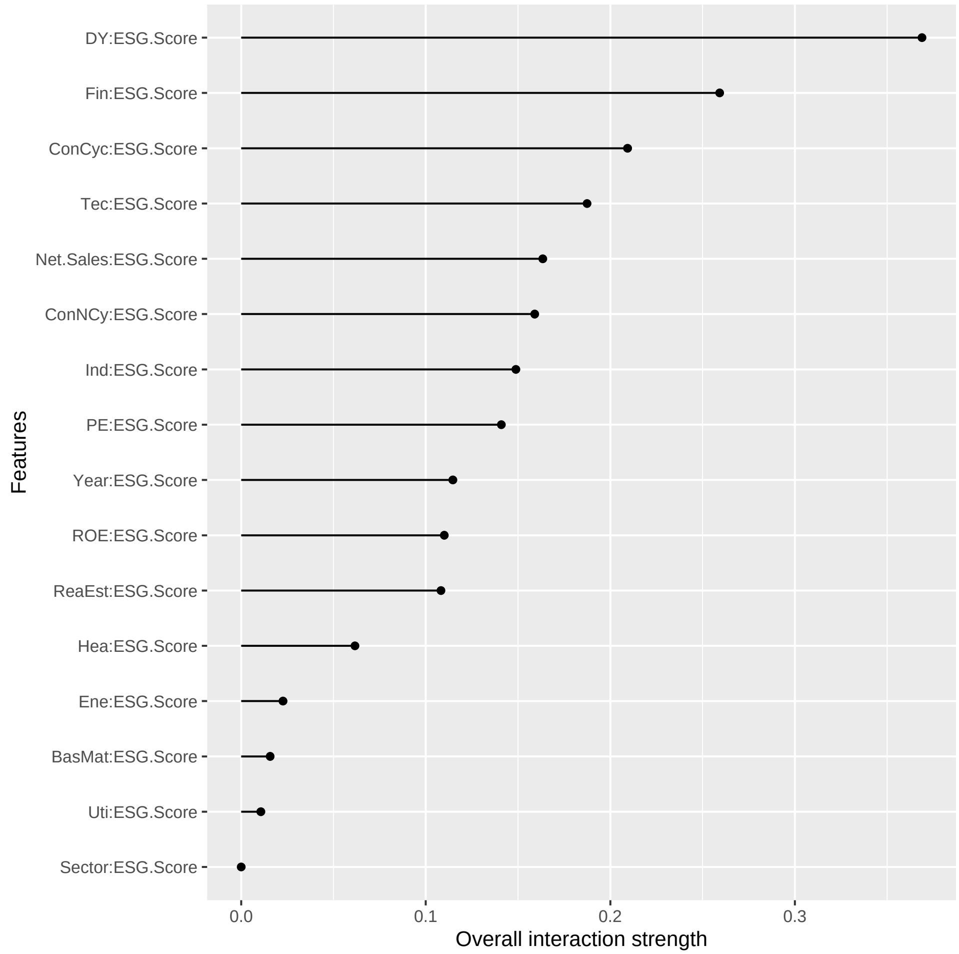
- *Overall, profiles are not parallel, indicating non-additive effects of the explanatory variable*
- *Part of the profiles suggests an approximately linear relationship between the ESG Score and the predicted EBIT value*

Interpretability: features interaction



Feature interaction - Each of the input features with all other features in predicting EBIT values.

Interpretability: features interaction



Two-way ESG:Score interactions with the other features in predicting EBIT values.

Interpretability: shapley value

- *Net.Sales* (= -8,190) decreases the prediction by 6,029
- *ESG.Score* (= 30.36) increases the prediction by 736
- *ROE* (= - 0.092) decreases the prediction by 700

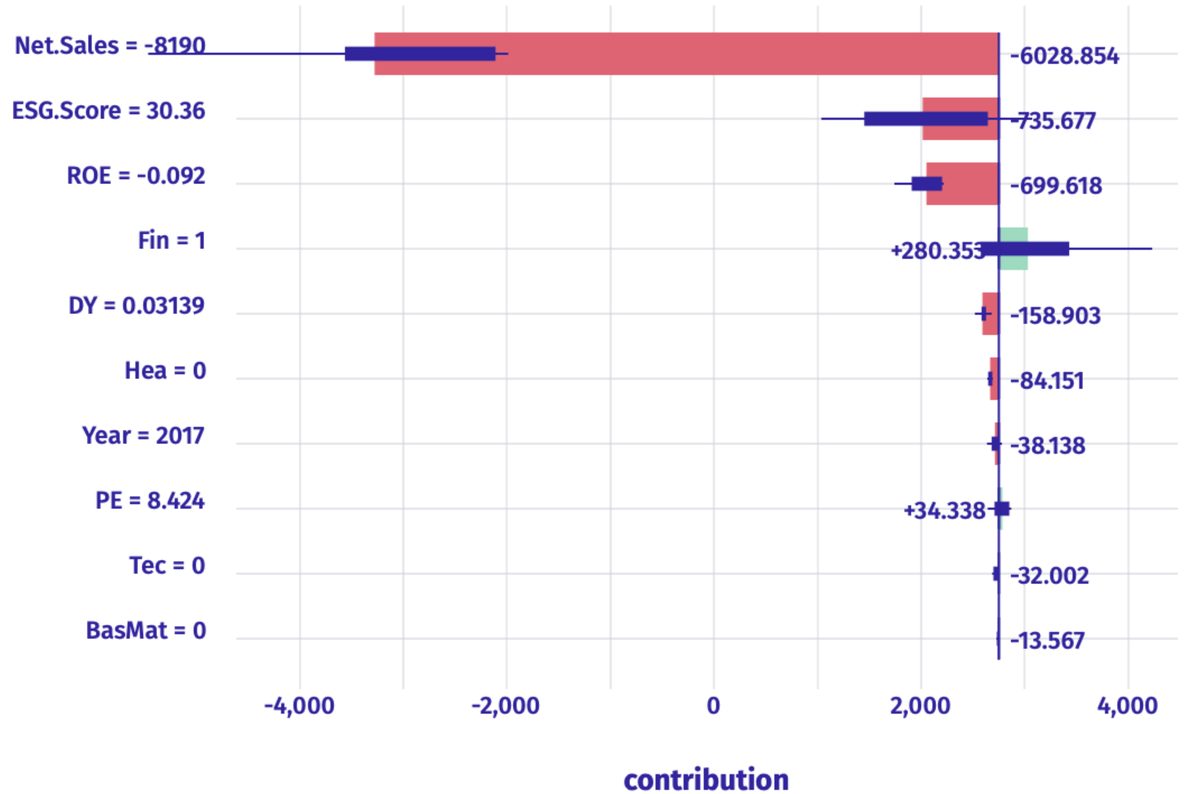


Figure: Shapley values. Data point: EBIT=-8,311. Red (green) bars show a negative (positive) contribution of the predictors.

Interpretability of the results: shapley value

- *ROE* (= 0.71) increases the prediction by 16,764
- *Net.Sales* (= 122,000) increases the prediction by 16,357
- *Hea* (= 1) decreases the prediction by 7,705

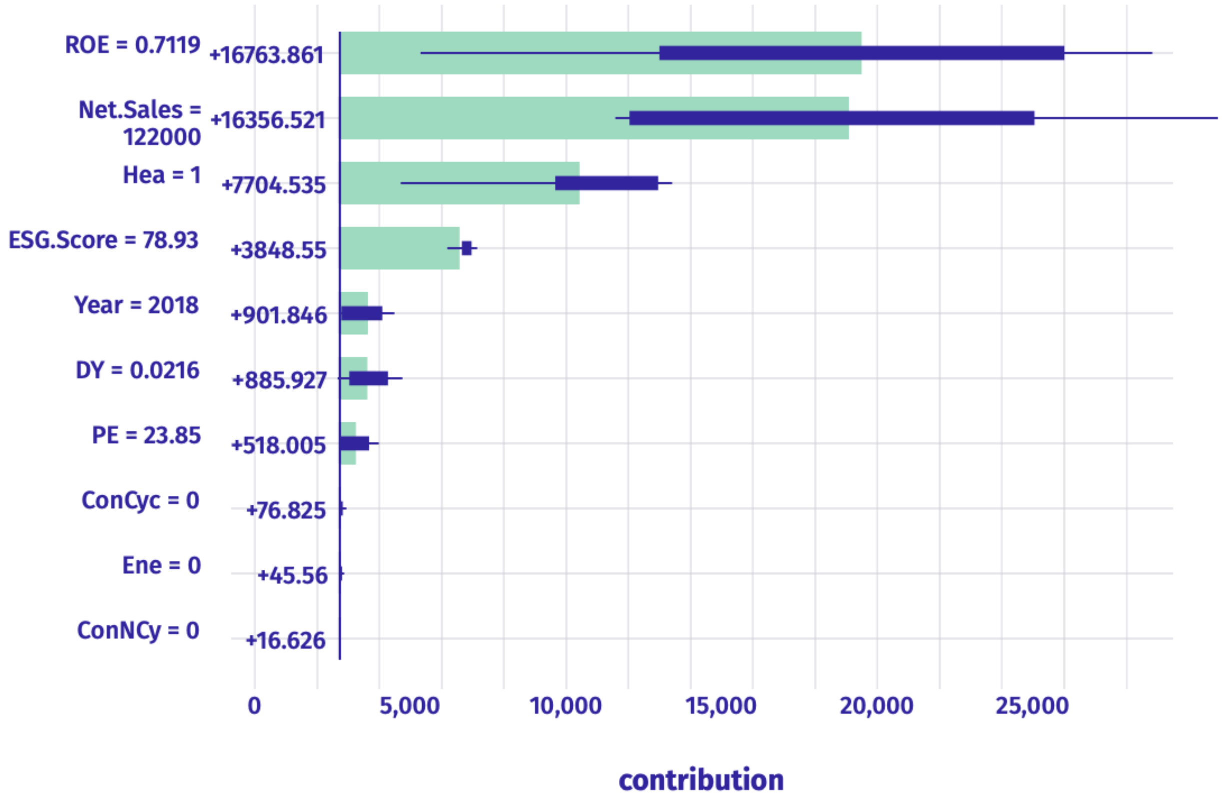
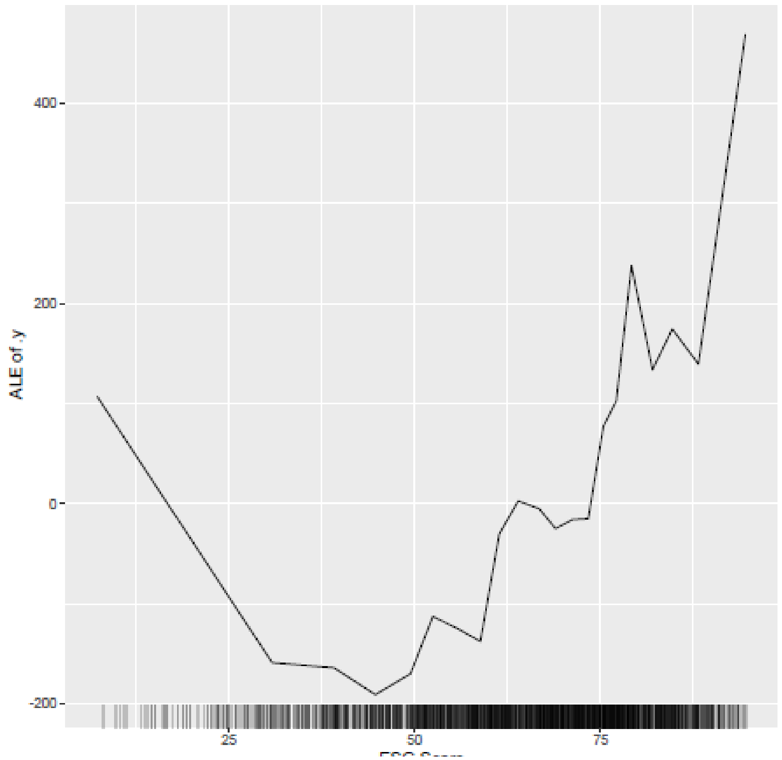


Figure: Shapley values. Data point: EBIT=53,683. Red (green) bars show a negative (positive) contribution of the predictors.

Interpretability of the results: ALE plot



- Marks on x-axis indicate the ESG score distribution
- Region 50-85 of the ESG score is the most relevant for the interpretation: the EBIT prediction rises with increasing ESG score
- Region 0-30 of the ESG score: the EBIT prediction decreases with increasing ESG score

Conclusions

- ESG score has a significant effect on the firm’s profitability

- Only a massive investment in sustainability and ESG criteria, which corresponds to higher ESG scores, leads to successful objectives, enhancing the strength of a company’s balance sheet

- Weak efforts in binding ESG elements into an investment strategy do not create an extra profit margin

Some References

1. Angel, J., Rivoli P. (1997). Does Ethical Investing Impose a Cost upon the Firm? A Theoretical Perspective, The Journal of Investing 6: 57–61.
2. Bauer, R., Koedijk, K., Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. Journal of Banking and Finance, 29(7):1751–1767.
3. Belghitar Y., Clark E., Deshmukh, N. (2014). Does it pay to be ethical? Evidence from the FTSE4Good. Journal of Banking and Finance, 47: 54–62.
4. Bello, Z.Y. (2005). Socially responsible investing and portfolio diversification. Journal of Financial Research, 28(1): 41–57.
5. Benabou R., Tirole J. (2010). Individual and Corporate Social Responsibility. Economica, 77.
6. Berg F., Koelbel J. F., Rigobon R. (2019). Aggregate Confusion: The Divergence of ESG Ratings, MIT Sloan Research Paper No. 5822-19.

- Bowen H. R. (1953). Social responsibilities of the businessman, Ethics and economics of society - New York, Harper.
- Breiman, L., Friedman, J. et al., Classification and regression trees. (1984)
- Breiman, L., Bagging predictors. Machine learning, 24(2):123–140. (1996)

Some References

1. Breiman, L., Random forests. *Machine learning* 45(1): 5–32. (2001)
2. Derwall J., Guenster N., Bauer R., Koedijk, K. (2005). The eco-efficiency premium puzzle. Financial Analysts Journal, 61(2): 51–63.
3. Hamilton S., Jo H., Statman, M. (1993). Doing well while doing good? The investment performance of socially responsible mutual funds. Financial Analysts Journal, 49(6): 62–66.
4. James, G., Witten, D., Hastie, T., Tibshirani, R., An Introduction to Statistical Learning: with Applications in R. Springer Texts in Statistics. ISBN 10: 1461471370. (2017)
5. Jewell, J., Livingston M. (1998), Split ratings, bond yields, and underwriter spreads. Journal of Financial Research, 21(2):185–204
6. Kreander, N., Gray R., Power D., Sinclair, C. (2005). Evaluating the performance of ethical and non-ethical funds: A matched pair analysis. Journal of Business Finance and Accounting, 32(7/8): 1465–1493.
7. Liaw, A., Package *randomforest*. https://cran.r-project.org/ web/packages/randomForest/randomForest.pdf. (2018)
8. Limkriangkrai, M., Koh, S., Durand, R.B. (2017). Environmental, Social, and Governance (ESG) Profiles, Stock Returns, and Financial Policy: Australian Evidence. International Review of Finance, 17(3): 461–471
9. Loh, W.Y., Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. (2011)
10. Quinlan, J.R., Induction of decision trees. Machine Learning, 1, 81–106. (1986)


감사합니다
Natick
Grazie
Danke
Ευχαριστίες
Dalu
Obrigado
Thank You
Köszönöm
Tack
Спасибо
Dank
Gracias
谢谢
Merci
Seé
ありがとう


it's
Q&A
TIME!