



POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE



EXECUTIVE SUMMARY OF THE THESIS

ESG rating construction: an objective and transparent approach

LAUREA MAGISTRALE IN MATHEMATICAL ENGINEERING - INGEGNERIA MATEMATICA

Author: CAMILLA CORTI, VERONICA CORTI

Advisor: PROF. FRANCESCA GRASSETTI

Academic year: 2021-2022

1. Introduction

Today's worldwide crises have been booming the interest for environmental, social and governance issues, shortly ESG issues. Since the valuation of the ESG corporate performance can influence financial decisions, with far-reaching impacts on corporate policies, customer preferences and asset prices, ESG rating agencies were born. Nevertheless, the lack of regulation on ESG disclosures generates poor and unreliable data collection, resulting in measurement inconsistency and subjective reporting, which make benchmarking very hard. In this regard, the Aggregate Confusion project by MIT Sloan Sustainability Initiative figures out a 0.61 correlation between ESG rating agencies¹, against the 0.99 between Moody's and Standard & Poor's credit ratings. It investigates the mismatches and identifies *scope*, *measurement* and *weights divergences* that is, what raters decide to measure, how they measure it, and the materiality they recognise[1]. Additionally, there is a lack of transparency about agencies' rating methodologies. These trends suggest a key challenge in putting all under a *common ESG len*.

We develop an algorithm which assesses a firm's ESG performance with an *objective* and *trans-*

parent snapshot of its impacts. To generate a common alphabet of disclosure, we rely on the Global Reporting Initiative Standards (GRIs). We split the analysis into the three environmental, social and governance pillars, finding pillars-specific key factors. Then, we group the GRIs, or KPIs, in a way that each group describes a key factor, integrating widely acknowledged norms, developed by the United Nations, the International Labor Organization, and the Organization for Economic Co-operation and Development. Finally, we measure each factor's impact on the overall score, relying on a complex research based on the UN Sustainable Development Goals (SDGs), specific literature and news sources. Our algorithm exploits a bottom-up approach and progressively develops a scoring process which aggregates a measurement and a weight component. Finally, to be as transparent as possible, it considers voluntarily data gaps via a penalty algorithm, and tries to fill them via predictive models.

At the end, we apply the algorithm on a set of Italian companies and we compare the results with Refinitiv and Sustainalytics ratings.

¹The project has focused on the rating agencies: KLD, Sustainalytics, Moody's ESG, S&P Global, Refinitiv, and MSCI.

2. ESG pillars

The section deeply deals with the analysis of key issues and topics populating the ESG scenario. In the following we summarise the discussion, presenting the common ESG blueprint of all the identified key factors valuing the company's long-term performance. We recall that each key factor collects a number of GRIs, Tables 1 and 3 are an example.

Environmental pillar
GHG, ODS and other significant emissions
Water
Land use and biodiversity
Raw materials sourcing
Waste and pollution
Clean-tech and renewables
Social pillar
Employment
Occupational and customer health and safety
Training and education
Modern slavery
Communities
Supplier social assessment
Product responsibility
Data privacy
Governance pillar
Economic performance and its impacts
Market presence
Business ethics

3. Constructing the rating

We develop an *objective* rating methodology which takes shape from businesses sustainable disclosures to think in a benchmarking scenario. The GRI documentation presents both qualitative and quantitative features to be disclosed. To remove subjectivity and be as *transparent* as possible, we translate the qualitative information into quantifiable data, wherever possible. In this way, the alphabet of disclosure collects original and revised GRIs, marked with a r . The algorithm builds a nested three-tiered structure based on convex linear combinations, whose final aim is the corporate ESG rating:

$$\varphi = \sum_{i \in \{e, s, g\}} \alpha p_i \quad (1)$$

where e , s , g are the environmental, social and governance pillars, $\alpha = 1/3$ and p_i is pillar i score, computed as:

$$p_i = \sum_{j=1}^{n_i} w_j \varsigma_j \quad (2)$$

where:

- n_i is the number of key factors in pillar i
- $(w_j)_{j \in \{1, 2, \dots, n_i\}}$ is the set of weights associated with the key factors of pillar i (see Section 4)
- $(\varsigma_j)_{j \in \{1, 2, \dots, n_i\}}$ is the set of scores associated with the key factors of pillar i

The core part of the procedure matches each key factor j to the score ς_j , via different and specific procedures which handle its KPIs. Precisely, it evaluates either plain indicators or their ratio. To ease the discussion, we refer to an "indicator".

3.1. Measurement tool

The algorithm learns the global environmental, social and corporate policies status, employing a training dataset. Thus, it processes a general rule for each indicator which is consistent with the current scenario. One at a time, it estimates the cumulative distribution function (CDF) which has generated each indicator sample points, by computing the corresponding empirical CDF (eCDF) which acts as a tool measuring the company relative performance.

We denote with K a generic indicator, taking values in D_K , with sample size m . To ease the notation, its realisations $(k_i)_{i \in \{1, 2, \dots, m\}}$ are assumed to be in ascending order. Since the set could have repetitions, we compute the highest number of recurrent elements:

$$n = \max_{i \in \{1, 2, \dots, m\}} n_i$$

where $n_i = \#\{j : k_j = k_i\}$. We define the eCDF $F_K : D_K \rightarrow [0, 1]$ as a step function jumping high by at most n/m at each observed point:

$$\begin{aligned} F_K(k) &= P(K \leq k) = \frac{1}{m} \sum_{i=1}^m \mathbb{1}_{(k_i \leq k)} \\ &= \frac{\#\{i : k_i \leq k\}}{m}, \quad k \in D_K \end{aligned} \quad (3)$$

We model the CDF of a discrete random variable since our training dataset is a finite collection of real data. However, since most of the variables take values in intervals of real numbers, we might model them as continuous random variables. Nevertheless, the difference between the step-wise CDF and any CDF f_K , obtained through a continuous interpolation, is at most equal to:

$$\epsilon = \|f_K - F_K\|_\infty = \max_{k \in D_K} |f_K(k) - F_K(k)| = \frac{n}{m}$$

Indeed, assuming m sufficiently large, and being $n \ll m$ reasonably, we get: $\lim_{m \rightarrow \infty} \epsilon = 0$, that is, as the sample size increases, the difference between the proposed solution and whatever kind of continuous CDF is negligible.

Evaluation process

Denoting with k the company registered value for K , K is mapped with a score s_K by evaluating F_K in k . Precisely, the procedure manages the indicator in this way:

- if K assesses any sustainable business practice moving the company long-term forward, then $s_K = F_K(k)$
- if K assesses a business practice somehow harming environmental protection, social justice or corporate accountability, then $s_K = 1 - F_K(k)$

Either way the algorithm values a non-negative contribute differing in the size of the recognised score. Since ratings have a meaning in comparison with the others, the contribute of a good indicator is grasped such that greater values correspond to greater scores, being the CDF a non-decreasing function. Contrarily, the procedure penalises misconducts delivering a score which is worth the less the more the size of the danger, being the opposite of the CDF a non-increasing function.

Overall we construct 52 indicators, defining the contribute of each of them step by step.

ESG score

Since it is a fair weighted sum of convex linear combinations in $[0,1]$, the comprehensive ESG score falls into $[0,1]$, where 1 and 0 indicate excellent and laggard capacity to fulfill sustainable commitments, respectively.

In the following we show two examples of key factors' analyses. Throughout the discussion we have:

- K_i : i -th indicator
- $(K_{ij})_{j \in \{1,2,\dots,m_i\}}$: training dataset for K_i
- m_i : number of K_i 's observations in the training dataset
- n_i : highest number of recurrent elements in $(K_{ij})_{j \in \{1,2,\dots,m_i\}}$
- k_i : value of K_i recorded by the company to be evaluated

Before proceeding, we perform some transformations on the dataset. The strict connection between raw materials, industrial products and

byproducts stands up from the resources consumption up to the waste generated by the industrial activity, making KPIs more meaningful when linked to the value of the production. Thus, in the environmental pillar, the dataset is divided by the corresponding total economic value generated. This transformation is denoted with a tilde.

Environmental pillar, Waste and pollution

rGRI	rKPI	symbol
306-2ar	Waste prevention	r
306-3a	Waste generated	x
301-4a	Waste diverted from disposal	y
301-5a	Waste directed to disposal	z

Table 1: Waste and pollution rKPIs, in tonnes.

The European Union Waste Framework Directive defines hierarchical guidelines for handling components end of life, including prevention, diversion from disposal and disposal. Accordingly, we combine the evaluations of the waste actually generated and its management practices, then we apply a reward function taking care of the prevention efforts².

Waste generated: dealing with business impact through waste generation, we define: $K_1 = \tilde{x}$ and we compute: $s_1 = 1 - F_{K_1}(k_1)$.

Waste management: the total amount of waste can be differentiated into diverted from and directed to disposal. We define: $K_2 = y/x$ and $K_3 = z/x$ and we compute $s_2 = F_{K_2}(k_2)$ and $s_3 = 1 - F_{K_3}(k_3)$. Finally, we combine the waste management practices computing the weighted sum of the evaluations of K_2 and K_3 , to get a comprehensive assessment:

$$s_{2,3} = k_2 F_{K_2}(k_2) + k_3 (1 - F_{K_3}(k_3))$$

Since $K_2 + K_3 = 1$, we have:

$$\begin{aligned} F_{K_2}(k_2) &= P(K_2 \leq k_2) = 1 - P(K_2 > k_2) \\ &= 1 - P(1 - K_3 > 1 - k_3) \\ &= 1 - P(K_3 < k_3) \end{aligned}$$

thus

$$F_{K_2}(k_2) = 1 - F_{K_3}(k_3)$$

²The reduction of some dangerous environmental impacts is always considered both in the relative key factor and in Clean-tech and renewables. We clarify it with an example. The rGRI 306-2ar Waste prevention is placed both in Waste and pollution and Clean-tech and renewables. However, we consider it in two different ways. In Clean-tech and renewables its contribute is computed through the usual CDF-reasoning, while, in Waste and pollution, it is applied a reward function to the overall key factor's score.

$$\forall k_3 \in D_{K_3}, k_3 \neq (K_{3j})_{j \in \{1,2,\dots,m_3\}}$$

where the difference is at most n_3/m_3 . Reasonably, $s_{2,3} = F_{K_2}(k_2)$. Thus, we combine the waste generated and the waste management with a fair convex linear combination:

$$s = 1/2 s_1 + 1/2 s_{2,3} \quad (4)$$

Waste reduction: focusing on the attempts to reduce waste, we define $K_4 = \tilde{r}$. We apply an adjusted function to the score created, which accounts for the size of the waste prevented, making use of the quantile function $Q_{K_4} : [0, 1] \rightarrow D_{K_4}$ which takes the value of the probability p and assigns x such that:

$$Q_{K_4}(p) = \max\{x \in D_{K_4} : P(K_4 \leq x) \leq p\}$$

We apply a continuous and increasing reward function to the computed rating (Eq.4) of the form $f_r(x) = (1+r)x$ where $r \in [0, 1]$ is the reward rate. Thus, we define the adjusted function $f_a(x) : [0, 1] \rightarrow [0, 1]$ such that $f_a(x) = \min(f_r(x), 1)$ to stop the function growth at 1. We compute $Q_{K_4}(1/3)$ and $Q_{K_4}(2/3)$ to split the reduction in these three awarding categories, matching them with a reward rate r :

Interval for k_4	r
$I_3 = (Q_{K_4}(2/3), 1]$	0.10
$I_2 = (Q_{K_4}(1/3), Q_{K_4}(2/3)]$	0.05
$I_1 = [0, Q_{K_4}(1/3)]$	-

Table 2: Waste prevention chart.

Finally, k_4 falls into one of these three ranges and the procedure computes the final score as $f_a(s)$.

Social pillar, Employment

Governance bodies at the start of the period		
rGRI	rKPI	Symbol
405-1ar.i	#men	$g_{m,b}$
405-1ar.ii	#women	$g_{w,b}$
405-1ar.v	#individuals under 30 years	$g_b^{(30)}$
405-1ar.vi	#individuals 30-50 years	$g_b^{(30-50)}$
405-1ar.vii	#individuals over 50 years	$g_b^{(50)}$
Governance bodies at the end of the period		
rGRI	rKPI	Symbol
405-1br.i	#men	$g_{m,e}$
405-1br.ii	#women	$g_{w,e}$
405-1br.v	#individuals under 30 years	$g_e^{(30)}$
405-1br.vi	#individuals 30-50 years	$g_e^{(30-50)}$
405-1br.vii	#individuals over 50 years	$g_e^{(50)}$
405-2ar	Remuneration of women to men	r

Table 3: Employment rKPIs. rKPIs concerning the employee group are omitted, they are exactly the same listing as the governance bodies' one.

We measure the company's level of diversity and inclusiveness in favour of equal opportunities, examining the composition of the workforce by age and gender, and his pay.

Distinction by gender: to discuss the difference in the gender composition at time t , we introduce:

$$\Lambda_{g,t} = \frac{|g_{m,t} - g_{w,t}|}{g_{m,t} + g_{w,t}}$$

We define:

$$K_1 = \Lambda_{g,e}$$

$$K_2 = \frac{\Lambda_{g,b} - \Lambda_{g,e}}{\Lambda_{g,b} \mathbb{1}_{(\Lambda_{g,b} > \Lambda_{g,e})} + (1 - \Lambda_{g,b}) \mathbb{1}_{(\Lambda_{g,b} \leq \Lambda_{g,e})}}$$

This disclosure provides a quantitative measure of the organisational diversity by computing $s_1 = 1 - F_{K_1}(k_1)$, combined with the evaluation of the company re-balancing efforts during the reporting period $s_2 = F_{K_2}(k_2)$. More precisely, K_2 quantifies the change plans implemented, reasoning as follows:

- if $\Lambda_{g,b} > \Lambda_{g,e}$ the gender composition of the governance body is somehow unbalanced at the beginning of the reporting period, but it is getting more gender balanced. This scenario is matched to the size of the variation over the maximum achievable improvement.
- If $\Lambda_{g,b} < \Lambda_{g,e}$ the gender composition of the governance body is getting more gender unbalanced during the reporting period. This scenario is matched to the size of the variation over the maximum possible worsening.
- If $\Lambda_{g,b} = \Lambda_{g,e}$ the gender composition of the governance body remains unchanged in the reporting period. This scenario is matched to zero to identify a middle situation in which neither an improvement nor a worsening has been made.

We resolve the governance body analysis by gender computing the convex linear combination of s_1 and s_2 , matching s_1 to the weight function $w_1 : [0, 1]^2 \rightarrow [1/2, 1]$ assuming this form:

$$w_1(x, y) = \frac{1}{2}(e^{-\alpha\sqrt{x^2+y^2}} + 1) \quad (5)$$

where $\alpha \in R^+$. Indeed, basing on the initial composition, as well as on the size of the variation in the reporting period, we recognise a different relevance to K_1 and K_2 , evaluating Eq.5 in $x = \Lambda_{g,b}$ and $y = |\Lambda_{g,b} - \Lambda_{g,e}|$.

Translating our reasoning into inequalities which define half-planes, we choose the value of α proceeding as follows:

- if $x \simeq 0 \wedge y \simeq 0$ the initial and the end compositions are close to the perfect balance. We prevent underestimating changes in the right neighborhood of zero, by strengthening the contribute of the final composition rather than the variation efforts:

$$w_1(x, y) > 0.9 \quad \forall x \leq 0.05 \wedge \forall y \leq 0.05 \quad (6)$$

- if $x \gg 0$ the initial composition is strongly unbalanced, making change plans desirable. Since there is no reason to increase one or other of the indicators, we fairly weight them:

$$w_1(x, y) < 0.55 \quad \forall x \geq 0.55 \wedge \forall y \quad (7)$$

To get a set of values satisfying Eqs.6 and 7, the mathematical model requires at least an upper bound of 0.55. Note that Eq.5 never takes exactly 1/2, but it approaches the value in the neighborhood of (1,1).

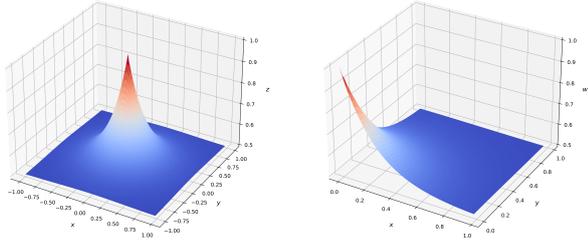


Figure 1: The left plot is the graph of the function with $\alpha = 4$ when $(x, y) \in [-1, 1]^2$. The right one is the graph restricted to the region of admissible values, i.e. $w_1(x, y)$ with $(x, y) \in [0, 1]^2$.

We get $\alpha \in [3.16, 4.19]$. Choosing $\alpha = 4$, we fix: $\bar{w}_1 = w(x, y)|_{x=\Lambda_{g,b}, y=|\Lambda_{g,b}-\Lambda_{g,e}|}$ and we finally compute the score:

$$s_{1,2} = \bar{w}_1 s_1 + (1 - \bar{w}_1) s_2 \quad (8)$$

Distinction by age: the analysis mirrors the previous one while focusing on the distinction by age. To investigate the distribution of the governance body in the three different categories (see Table 3) at time t , we introduce two quantities resembling the unbiased estimators of a population mean and variance. We imagine of fairly dividing the individuals in the three different classes, defining the pseudo-mean:

$$\mu_{g,t} = \frac{g_t^{(30)} + g_t^{(30-50)} + g_t^{(50)}}{3}$$

and we measure the dispersion of the categories'

values with respect to the fair distribution of the individuals in the three classes with the pseudo-variance:

$$\sigma_{g,t}^2 = \frac{(g_t^{(30)} - \mu_t)^2 + (g_t^{(30-50)} - \mu_t)^2 + (g_t^{(50)} - \mu_t)^2}{2}$$

We finally introduce:

$$\Gamma_{g,t} = \sigma_{g,t}^2 \quad (9)$$

which quantifies the individuals' scatter in the different age groups with respect to the total number of individuals. Thus, we define:

$$K_3 = \Gamma_{g,e}$$

$$K_4 = \frac{\Gamma_{g,b} - \Gamma_{g,e}}{\Gamma_{g,b} \mathbb{1}_{(\Gamma_{g,b} > \Gamma_{g,e})} + (\sigma_{max\ g,b}^2 - \Gamma_{g,b}) \mathbb{1}_{(\Gamma_{g,b} \leq \Gamma_{g,e})}}$$

where $\sigma_{max\ g,b}^2$ is the maximum of Eq.9 at the beginning of the reporting period.

Computation of $\sigma_{max\ g,t}^2$ we fix $x = g_t^{(30)}$, $y = g_t^{(30-50)}$, $z = g_t^{(50)}$ and $N = x + y + z$. Precisely, being the number of individuals within the organisation's governance bodies in a specific age category, x , y and z can assume only positive or null values. We compute the maximum solving a constrained maximisation problem of a function in three variables, exploiting the Lagrange multiplier method. We define:

$$L(x, y, z, \lambda) = f(x, y, z) - \lambda g(x, y, z)$$

where:

$$f(x, y, z) = \frac{(x - \frac{N}{3})^2 + (y - \frac{N}{3})^2 + (z - \frac{N}{3})^2}{2}$$

$$g(x, y, z) = x + y + z - N$$

where $f(x, y, z)$ is the pseudo-variance and $g(x, y, z)$ is the constraint to which f is subjected. Then, we solve the system:

$$\begin{cases} \frac{\partial L(x,y,z,\lambda)}{\partial x} = 0 \\ \frac{\partial L(x,y,z,\lambda)}{\partial y} = 0 \\ \frac{\partial L(x,y,z,\lambda)}{\partial z} = 0 \\ \frac{\partial L(x,y,z,\lambda)}{\partial \lambda} = 0 \end{cases} \rightarrow \begin{cases} (x - \frac{N}{3}) - \lambda = 0 \\ (y - \frac{N}{3}) - \lambda = 0 \\ (z - \frac{N}{3}) - \lambda = 0 \\ x + y + z - N = 0 \end{cases}$$

yielding to $\lambda = 0$ and to the stationary point $(x, y, z) = (\frac{N}{3}, \frac{N}{3}, \frac{N}{3})$. Precisely, since $f(x, y, z)$ is non-negative and $f(\frac{N}{3}, \frac{N}{3}, \frac{N}{3}) = 0$, $(\frac{N}{3}, \frac{N}{3}, \frac{N}{3})$ is a minimum point for f . Given that we have only one stationary point and the domain of the function is compact, there must be a maximum. The maximum is on the extremes of the domain,

which is either $x = 0$ or $y = 0$ or $z = 0$. Being f symmetric in the three variables, we fix $z = 0$. Thus, $y = N - x$ and substituting the values in f , we get:

$$f(x, y) = x^2 - Nx + \frac{N^2}{3}$$

which is the equation of a concave upward parabola, whose minimum corresponds to $(x, y) = (\frac{N}{2}, \frac{N}{2})$. Finally, we find the maximum of $f(x, y)$ at the extremes of the interval, either in $x = 0$ or $x = N$. Replacing the value obtained in f we get the maximum, that is $\sigma_{\max, t}^2 = N^2/3$.

Reasoning as before, we weight K_3 and K_4 exploiting Eq.5 with $x = \Gamma_{g,b}$, $y = |\Gamma_{g,b} - \Gamma_{g,e}|$ and $\alpha = 4$:

$$s_{3,4} = \bar{w}_3 s_3 + (1 - \bar{w}_3) s_4 \quad (10)$$

Exactly the same analysis is computed for the employee group, defining the indicators K_5 , K_6 , K_7 and K_8 and computing the scores $s_{5,6}$ and $s_{7,8}$.

Pay: we measure the remuneration of women to men defining:

$$K_9 = \max(r, r^{-1})$$

Since the fraction could be both greater and smaller than one, to synthesise a criterion penalising the more values deviate from the unity, in favour of equal pay, we select the maximum and we compute:

$$s_9 = 1 - F_{K_9}(k_9) \quad (11)$$

Final score: since leadership roles require some experience, they are typically played by older individuals. Indeed, we recognise to all the scores (Eq.8, $s_{5,6}$, $s_{7,8}$ and Eq.11) the same contribute, $w = 0.225$ but Eq.10, which is matched to $w = 0.10$:

$$s = 0.225 s_{1,2} + 0.10 s_{3,4} + 0.225 s_{5,6} + 0.225 s_{7,8} + 0.225 s_9$$

3.2. Penalty

Missing values plague ESG dataset, making it challenging to draw up accurate evaluations. For example, when a company communicates a little amount of ESG data and this indicates a good long-term performance, ending up with a high ESG score would not be fair neither honest. Moreover, if the lack of data is explicitly considered, companies would be encouraged to

provide new additional information. Considering only the original GRIs – and not those revised – we define a ESG datum gap when it is demanded between the GRI Standards but the company does not appropriately disclose it. Denoting with m_i the number of data gaps in pillar i , $i \in \{e, s, g\}$, the algorithm matches any datum hole to a score reduction of α_i times the comprehensive score of the pillar to which it belongs. Precisely, when the total number of gaps accounts for half of the required data (or its ceiling function, if it is not an integer number), the procedure reduces by 1/4 a before-penalty score of 1.

$$\left\lceil \frac{m_i}{2} \right\rceil \alpha_i = \frac{1}{4} \quad (12)$$

In this way, each pillar has its own value for α depending on the grade of misinformation a datum gap may represent.

4. Weights

In this section we select the weights competing with each key factor, to measure their impact on the overall score. It is a complex capillary research based on SDGs, specific literature and ESG news sources coming from publicly available information. Since each pillar follows a specific reasoning, we split the analysis in three sections. In general, with the exception of the governance pillar, each weight combines the information about how the relative industry contributes to the main concern in comparison with the other sectors (e.g. how emission-intensive the specific industry is with respect to the others), with the time frame in which the issue may materialise.

Recalling Eq.2, key factors weights are chosen such that:

$$\begin{cases} 0 \leq w_j \leq 1 \\ \sum_{j=1}^{n_i} w_j = 1 \end{cases}$$

Environmental weights

According to the Global Industry Classification Standard, we classify the existing industries into 11 industrial sectors³.

The environmental methodology identifies an

³The Global Industry Classification Standard is the industrial taxonomy developed by MSCI and S&P Global, which includes: Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Communication Services, Utilities and Real Estate.

industrial-specific weighting system since, basing on the nature of its activities, each firm has its own environmental risks and opportunities. For example, more than 2/3 of worldwide greenhouse gas emissions come from the energy sector[2], while they are estimated to be 21-37% for the food industry[3]. Even in evaluating the same ESG choice, we develop different assessment criteria following this guideline: if the considered issue strictly deals with the business main activity, this features a high weight, otherwise a lower one. In this way, we standardise the efforts of different companies in different industries, allowing for a fairer relative comparison.

Social weights

The social dimension is not as much closely related to the company’s main activity, as to its operations’ effectiveness in affecting and shaping people’s and communities’ labour and welfare conditions. Accordingly, all companies should embrace human rights due diligence to identify, prevent and mitigate human rights impacts. However, even if in general the impacts of social issues encompass the various market sectors, for certain industries they are even more significant than others. We select only two weights types, distinguishing more significant from base factors. We find out the industrial sectors being more responsible for some key factors, implementing one, two or at most three, out of eight, further risky social actions, relying on the existing documentation. In this way, we classify the industrial sectors into three different categories k_i , $i \in \{1, 2, 3\}$, basing on the number i of their most significant issues. Indicating with x_i the base weight of a generic factor in category i , we size the materiality of the further significant issues doubling their weights, i.e. $y_i = 2x_i$. Thus, x_i and y_i are the solutions of this system of two equations in two variables:

$$\begin{cases} y_i = 2x_i \\ 1 = (8 - i)x_i + iy_i \end{cases} \rightarrow x_i = \frac{1}{8+i}$$

getting: $x_1 = 1/9$, $x_2 = 1/10$ and $x_3 = 1/11$.

Governance weights

The complex network of figures surrounding any type of company requires an entity running the enterprise, which is responsible for the management of the entire conglomerate. Indeed, the governance weights are set equal between all industries. Considering the breakdown of corpo-

rate governance issues in the perspective of a sustainable performance analysis, we overlook the key factor tax since taxes payment is unavoidable.

5. Data prediction

We exploit data transformation handling logarithms to deal with large ranges and skewed distributions[4][5][6].

We use linear regression predictive models on log-log transformed data.

Training dataset We collect the training dataset from Refinitiv Eikon. The baseline is 2020.

Definition 5.1. The Weighted Least Square regression computes the regression coefficients minimising the weighted mean square error:

$$\min_{(q,m) \in \mathbb{R}^2} \frac{1}{N} \sum_{i=1}^N w_i (y_i - q - mx_i)^2 \quad (13)$$

where:

- $(x_i, y_i)_{i \in \{1,2,\dots,N\}}$ are the observations
- q is the intercept and m is the regression coefficient
- $w \in \mathbb{R}^N$ are the observations’ weights

Fixing $w_i = 1$, $\forall i \in \{1, 2, \dots, N\}$ in Eq.13, we get the Ordinary Least Square regression coefficients. We test the extended least square assumptions, exploiting the Breusch-Pagan test to fix any linear form of heteroskedasticity. When dealing with heteroskedasticity, we re-compute the regression coefficients using Eq.13, estimating w as the inverse of the residuals variance, the inverse of the predicted value of y and the inverse of the square predicted value of y . In the following, we present two of our results:

Direct (Scope 1) GHG emissions to NO _x , SO _x and other significant air emissions			
Model	Fitted coefficients		Statistic
	q	m	R^2
WLS - $w = 1/\hat{y}^2$	7.8579	0.7903	0.6920
Water discharge to water withdrawal			
Model	Fitted coefficients		Statistic
	q	m	R^2
OLS	0.5345	0.9206	0.7040

Table 4: Selected predictive models for the regression of X to Y.

6. Computing the rating

Dataset to be evaluated

To test our evaluation criterion, we manually collect companies' sustainability data. These data concern 222 Italian listed companies in the FTSE All Share index (the baseline is 2020, as before).

Comparison with the ratings of Refinitiv and Sustainalytics

We make use of data coming from two of the major ESG rating providers, Sustainalytics and Refinitiv, manually collecting Sustainalytics evaluations from Yahoo Finance and Refinitiv ones from Refinitiv Eikon. To make comparisons, we apply some adjustments to the raw data: first, we only take into account rating scores submitted at the end of the year, second, a scale ranging from 0 to 1 is exploited to translate different rating scores given by various agencies. When comparing different ESG rating providers evaluations, we distinguish two different cases:

- *data samples as a snapshot of reality*: accepting that data samples are a valuable representation of reality, we can make absolute considerations between the three rating agencies different assessments.
- *Non uniform data distribution in the range of acceptable values*: relaxing the assumption of uniform distribution of data, we can only make claims which are relative to the maximum value reached by the sample.

Even if the analysis is constrained by data availability, we confirm previously cited assessments on ESG rating divergence. Precisely, we get more discrepancies in the environmental pillar evaluations than in the social ones. Indeed, our algorithm issues an average environmental ESG rating score of 0.3511 while Refinitiv associates 0.5713 and Sustainalytics associates 0.6802. Moreover, the analysis of the results shows an important point: prediction models are a key tool for taking advantage of all the information provided by sustainability reports, however, they should be handled cautiously. Unsuitable regression curve estimates may alter scoring results.

References

- [1] Florian Berg, Julian F Koelbel, and Roberto Rigobon. Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6):1315–1344, 2022.
- [2] European Council. 5 facts about the EU's goal of climate neutrality, 2021. URL www.consilium.europa.eu/en/5-facts-eu-climate-neutrality/.
- [3] Priyadarshi R Shukla, J Skeg, E Calvo Buendia, Valérie Masson-Delmotte, H-O Pörtner, DC Roberts, Panmao Zhai, Raphael Slade, Sarah Connors, S Van Diemen, et al. Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. 2019.
- [4] Douglas S Glazier. Log-transformation is useful for examining proportional relationships in allometric scaling. *Journal of theoretical biology*, 334:200–203, 2013.
- [5] Paul M Meaney, Qianqian Fang, Tonny Rubaek, Eugene Demidenko, and Keith D Paulsen. Log transformation benefits parameter estimation in microwave tomographic imaging. *Medical physics*, 34(6Part1):2014–2023, 2007.
- [6] Isabella Zwiener, Barbara Frisch, and Harald Binder. Transforming RNA-Seq data to improve the performance of prognostic gene signatures. *PloS one*, 9(1):e85150, 2014.