Ca' Foscari University of Venice

Department of Economics

Working Paper

Diana Barro
Marco Corazza
Gianni Filograsso

A ESG rating model for European SMEs using multi-criteria decision aiding

ISSN: 1827-3580
No. 27/WP/2023

Working Papers
Department of Economics
Ca' Foscari University of Venice
No. 27/WP/2023
ISSN 1827-3580

A ESG rating model for European SMEs using multi-criteria decision aiding

Diana Barro
Ca' Foscari University of Venice

Marco Corazza
Ca' Foscari University of Venice

Gianni Filograsso
Ca' Foscari University of Venice

Abstract
Through ESG assessment, companies can effectively measure their exposure to environmental, social, and governance (ESG) risks identifying opportunities for long-term sustainable growth and future social and environmental impact. This process is crucial for listed small and medium-sized enterprises (SMEs) wanting additional support in their ESG transition. The importance of such assessments will only intensify in the future as the implementation of the Sustainable Finance Disclosure Regulation (SFRD) and the Corporate Sustainability Reporting Directive (CSRD) will require all listed companies to be on equal footing. In this contribution, we propose to apply a multi-criteria method (MURAME) to assess the sustainability profiles of SMEs. The methodology, which allows for measuring a firm's environmental, social, and governance (ESG) efforts, is applied to a sample of European-listed SMEs with the aim of identifying ESG leaders and laggards and analyzing potential sector-specific effects. The obtained ranking results show some degree of robustness across different model parameterizations. Furthermore, we propose to model the benefits of voluntary disclosure of sustainability information under a prudential scoring framework.

Keywords
corporate social responsibility, sustainability policy, small- and medium-sized enterprises, multi-criteria decision making

JEL Codes
C44, Q56, M14, O16

Address for correspondence:
Diana Barro
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
e-mail: d.barro@unive.it

The Working Paper Series
is available only on line
(http://www.unive.it/pag/16882/)
For editorial correspondence, please contact:
wp.dse@unive.it

Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta San Giobbe
30121 Venice Italy
Fax: ++39 041 2349210

A ESG rating model for European SMEs using multi-criteria decision aiding

Diana Barro* Marco Corazza* Gianni Filograsso*

October 27, 2023

Abstract

Through ESG assessment, companies can effectively measure their exposure to environmental, social, and governance (ESG) risks identifying opportunities for long-term sustainable growth and future social and environmental impact. This process is crucial for listed small and medium-sized enterprises (SMEs) wanting additional support in their ESG transition. The importance of such assessments will only intensify in the future as the implementation of the Sustainable Finance Disclosure Regulation (SFRD) and the Corporate Sustainability Reporting Directive (CSRD) will require all listed companies to be on equal footing regarding ESG reporting. In this contribution, we propose to apply a multi-criteria method (MURAME) to assess the sustainability profiles of SMEs. The methodology, which allows for measuring a firm's environmental, social, and governance (ESG) efforts, is applied to a sample of European-listed SMEs with the aim of identifying ESG leaders and laggards and analyzing potential sector-specific effects. The obtained ranking results show some degree of robustness across different model parameterizations. Furthermore, we propose to model the benefits of voluntary disclosure of sustainability information under a prudential scoring framework.

Keywords: Corporate social responsibility, Sustainable policy, Small and medium-sized enterprises, Multi-criteria decision aiding

JEL Codes: C44, Q56, M14, O16

*Dept. of Economics, Ca' Foscari University of Venice

1 Introduction

The sustainability assessment of a company is becoming a critical issue for European SMEs, thanks to pressure from involved stakeholders, such as regulators, competitors, clients, NGOs, media, shareholders, and employees (Torelli et al., 2020). Moreover, the Corporate Sustainability Reporting Directive (CSRD), entered into force in January 2023, sets the bar higher in terms of transparency and requirements and shifts the focus from non-financial to sustainability reporting (Baumüller and Grbenic, 2021), compared to the previous Non-Financial Reporting Directive (NFRD), including a broad set of new topics to be covered. The directive extends the scope of companies required to report on sustainability topics from 11,700 to approximately 50,000 (European Commission, 2023).

More precisely, the new directive calls for a more comprehensive report on the impact of corporate activities on the environment and society and mandates independent audits of reported information. The resilience of a company to related risks must be taken into account as well, provided they are *material* to the business activities. Materiality can be defined as *"the potential adverse effects on system elements significantly impairing the performance of the economic activity"* (Deloitte, 2023). Social and governance metrics must be reported along with environmental performance, including -among others- human rights, anti-corruption, and diversity across management.

In this study, we aim to contribute to corporate risk management in SMEs literature by examining ESG profiles, whose increasing importance is having a tangible effect on companies' overall financial health (Höck et al., 2020). In particular, traditional credit assessments focus on the competitive position and the corporate and country risk, along with the cash flow, leverage, and governance evaluation (Siddiqi, 2017). However, each of the above-mentioned topics can be affected by environmental, social, and governance factors. Therefore, a *double materiality* assessment, i.e. how the company's business and outlook is affected by sustainability risks and how the company's activities directly or indirectly impact the society and the environment, is of utmost importance.

First, the analysis of the environmental pillar addresses how the environmental impact of a business can affect its risk profile. There are four main transmission channels to consider: climate risk, transition risk, regulatory risk, and reputational risk. The analysis should entail how the natural resources used by a firm -directly or indirectly- can increase exposure to such risks, ultimately hindering the firm's financial stability.

Second, the social pillar addresses the relationships and interactions of a firm with all its stakeholders and, more broadly, with society. From a double materiality perspective, therefore, it entails an analysis of safety and human capital management, social cohesion, support for diversity, and reduction of inequalities.

Finally, the governance pillar addresses how the arrangement of rules and processes by which a firm is managed can influence its risk profile through operational, legal, or reputational risk transmission channels. Broadly speaking, it requires assessing whether the interests of the company are effectively managed through a system of checks and balances, with particular attention to the relationship with the stakeholders. Lack of independence and diversity in a firm's board, along with poor transparency and planning, provide early warning signals of deteriorating financial health.

(Giese et al., 2019), in line with the previous seminal contribution of (El Ghoul et al., 2011), identify three transmission channel channels between ESG performance and valuation within a standard discounted cash flow (DCF) model, namely the cash-flow channel, the idiosyncratic risk channel and the valuation channel. According to the authors' analysis, under relatively strict assumptions of no direct and indirect costs for ESG disclosure (Prencipe, 2004), the ESG information is transmitted to the company's valuation and performance through lower cost of capital, higher profitability, and lower tail risk. Their assessment is limited to large-cap companies.

Previous attempts to capture the ESG performance of large-cap firms have also been made with textual analysis (Baier et al., 2020).

Availability of data constitutes a critical issue for ESG modeling (Kotsantonis and Serafeim, 2019), even more for SMEs, likely due to limited benefits in disclosing sustainability information (Gjergji et al., 2021). Such an issue has also been raised for the credit rating of SMEs (Angilella and Mazzù, 2015), for which firms have clear incentives to release data publicly.

The modelization of the SMEs' ESG performance, along with a comprehensive assessment of major implications for firms and policymakers, despite its importance, especially in the European context, has not received enough attention in the literature on financial risk management in SMEs.

Previous studies on financial risk management in SMEs have mainly focused on the credit risk channel for the transmission of shocks. Quantitative techniques have been widely applied to model and manage credit risk in SMEs, including statistical methods (Altman and Sabato, 2007; Altman et al., 2010), fuzzy analysis (Roy and Shaw, 2021) and multi-criteria decision methods (MCDM) (Voulgaris et al., 2000; Angilella and Mazzù, 2015; Corazza et al., 2016).

In light of this research gap, we tackle this problem by deliberately putting an emphasis on interpretability and flexibility rather than complexity. Consequently, our aim is to marry the specific peculiarities of SMEs and the need to converge towards higher disclosure standards by proposing to construct a sustainability scoring model based on the well-established multicriteria ranking method (MURAME) (Goletsis et al., 2003) and to focus on actual implications of the model assumptions for firms and policymakers.

On the basis of a hand-collected dataset of European SMEs, we aim to identify top and bottom firms (leaders and laggards) and to capture potential sector-specific effects. Furthermore, we want to investigate to which extent, assuming a prudential scoring methodology, voluntary disclosure of ESG information is actually beneficial.

The rest of the paper is organized as follows. Section 2 briefly presents research trends in the literature on financial risk management and sustainability in SMEs. Section 3 presents how data has been collected and pre-processed, whereas Section 4 briefly presents the methodology. Section 5 is devoted to the presentation and discussion of the results. Finally, Section 6 concludes and presents some future research directions.

2 Literature review

The peculiar features of SMEs have drawn the interest of scholars and practitioners alike a long time ago (Voulgaris et al., 2000), especially with respect to the creditworthiness assessment problem (Doumpos and Figueira, 2019) since the credit supply is also among the primary transmission channels of economic shocks for SMEs (D'Amato, 2020). In the valuation of the issue of financing to such companies, the inclusion of non-financial information is potentially as crucial as standard financial ratios used for credit risk assessment. As (Altman et al., 2010) argue, SMEs require risk management tools and methodologies specifically tailored to their needs, and that can also be applied in the case of poor, unreliable, or simply lack of disclosure on crucial topics. A non-exhaustive list of relevant features includes operational risk, the type, size, age, and sector of the business.

A strand of the literature has, in particular, highlighted the necessity of developing specific credit risk models for innovative SMEs (Czarnitzki and Hottenrott, 2011), including a range of econometric and multicriteria decision aid models. In this case, lending relies necessarily on soft information (Moro and Fink, 2013).

More recently, the importance of assessing sustainability practices and sustainability performance in SMEs has emerged. In a comprehensive review, (Malesios et al., 2021) investigate the importance of the sustainability strategies and processes that firms employ to reduce environmental impact, enhance positive social impact, and create long-term value for stakeholders, aiming at achieving sustainable values. As for such topics, SMEs display unique features that must be taken into account for the construction of a scoring system.

Previous studies have emphasized the size of a company itself as one of the key drivers for the adoption of corporate social responsibility (CSR) standards (Goyal et al., 2013). As a result, the adoption of such practices has been previously highlighted as potentially detrimental in terms of the cost-benefits structure (Gjergji et al., 2021) for SMEs.

As far as the benefits are involved, compared to large corporations, the managerial structure of SMEs is simpler: agency costs are therefore smaller or missing, potentially because the agent and the principal coincide or direct supervision occurs (Bartolacci et al., 2020). From the perspective of costs, the picture is less clear. (Gjergji et al., 2021) argue that either direct or indirect economic costs are among the main barriers to the development of CSR in SMEs. Companies indeed bear direct costs due to required expertise and investments in sustainable reporting (production and dissemination costs) and indirect costs from disclosing segment information (competitive costs). Both are larger for SMEs (Prencipe, 2004), mainly because of the larger fixed component of costs and the difficulty for small firms in protecting from competitors. (Rodríguez-Gutiérrez et al., 2021), instead, claim that it is one of the least valued criteria, suggesting that there is widespread awareness among SMEs that CSR can yield long-term returns and that economic cost is a required precondition.

Nonetheless, the preparation of a sustainability report requires a decision-making process characterized by standard rules: lack of operational tools, along with technical knowledge and advertising skills (Gjergji et al., 2021) are major concerns for all SMEs that wish to formalize such sustainability practices.

To summarize, if, on the one hand, the impact of ESG disclosure is widely recognized as positive for large firms' valuation (El Ghoul et al., 2011), since voluntary non-financial information can effectively mitigate exposure to a broad range of direct and indirect risks, the relationship is way less clear for SMEs. The lack of clarity on the relationship between market valuation and ESG ratings of SMEs can be essentially boiled down to a different cost-benefit structure or to the potential mispricing of such effects.

Due to SMEs' scant track records, the Multiple Criteria Decision Aid (MCDA) approach (Zionts, 1979), thanks to its flexibility, is especially effective for such unstructured problems. For the credit rating models, a large volume of studies is available for both traditional and innovative SMEs (Voulgaris et al., 2000; Angilella and Mazzù, 2015; Corazza et al., 2015, 2016; Roy and Shaw, 2021). A strand of the literature has proposed using hybrid approaches, i.e. MCDA methods based on the combination of MCDA strategies. Several models have been proposed that integrate sustainability information to develop sustainable credit scoring for SMEs, mainly based on the Analytic Hierarchic Process (AHP) (Saaty, 1988), on the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) (Hwang et al., 1981) or on the Best-Worst Method (BWM) (Rezaei, 2015), and their respective fuzzy versions.

However, to the best of our knowledge, little attention has been paid to the construction of an ESG rating system for SMEs in the MCDA literature.

(García-Martínez et al., 2019) propose a combination of a goal programming approach and factor analysis for dimensionality reduction in order to assess the optimal combination of fundamental CSR score drivers of European large-cap firms.

(Guney et al., 2020) focus on the governance performance of large-cap companies according to the ASSET4 ESG module from Datastream and construct a competing corporate governance quality indicator through PROMETHEE II (Brans and Vincke, 1985). They find that it is less susceptible to the endogeneity issue and more robust across different subsamples, corporate performance indicators, and industries. Finally, they document a strong and negative relationship between corporate governance and corporate performance, which can be boiled down to the impact of associated costs.

Finally, (Rodríguez-Gutiérrez et al., 2021) propose a novel methodological approach, based on BWM, to understand which criteria are considered crucial for initiating sustainability reporting by Spanish SMEs. In this way, the authors aim at identifying the costs and the benefits that could arise from the adoption of sustainability reporting.

3 Data

In what follows, we illustrate in detail the construction and cleaning process of the hand-collected dataset used in the analysis. Starting with the screening of the sustainability reports for an initial sample of listed European SMEs we identified a set of relevant criteria, we dealt with ESG data quality issues, and we discussed a proposal for handling missing data.

3.1 Data collection and pre-processing

In this section, we discuss the feature extraction process and the pre-processing step to handle missing values. The primary source for identifying the perimeter of European listed SMEs was the Orbis database of Bureau van Dijk. In accordance with the definition established by the European Commission, we searched for active firms with less than 50 million turnover and less than 250 employees, then we hand collect data starting from a dataset of 1,337 listed European SMEs, retrieved on 26 April 2023. Practically speaking, this implies that most firms included in the sample are, according to standard practice and from a market capitalization point of view, microcap companies. From this set of companies, we collect all the relevant sustainability reports, which are available for around the 20% of the initial set of firms, and then we filter the data. All the criteria found to be compliant with (Global Reporting Initiative, 2022) standards are collected, then binary variables are removed. Firms reporting data for more than 50% of selected criteria are kept: as a result, we collect an overall number of 104 companies, therefore less than 10% of the initial sample.

Furthermore, data are normalized by sector in a range $[0, 1]$, in order to neutralize the impact of sector-specific features, which is a common practice in financial economics and in the ESG literature (Sorensen et al., 2021). However, the relatively small sample size allows us to derive only sector-neutral scores, whereas we leave to future work the refinement of such neutralization step at industry and country-level detail, as we expect to see more companies to disclose non-financial information in the future and rating agencies to extend the universe of ESG-rated firms, due to pressure from different stakeholder categories (Torelli et al., 2020). Better coverage and higher data quality would allow to explicitly incorporated sector and country effects into the model, in line also with recent contributions in the field of corporate default risk prediction (Doumpos et al., 2017).

Therefore, we perform a linear transformation on the original data, such that the scaled data are in the range $[0, 1]$, with j denoting the alternatives for a given criterion and $\theta \in$ {*Industrials, Consumer discretionary, Consumer staples, Health Care, Financials, Information Technology, Communication services, Utilities, Real Estate*} corresponds to the sector the alternative j belongs to:

$$X_{scaled,\theta j} = \frac{X_{\theta j} - X_{min,\theta j}}{X_{max,\theta j} - X_{min,\theta j}} \tag{1}$$

As for the missing data, there is a large volume of both theoretical (Tsikriktsis, 2005) and ESG-related (Kotsantonis and Serafeim, 2019; Sahin et al., 2022) contributions tackling this issue. However, note that the measurement of ESG ratings is quite ambiguous, and existing definitions are sometimes competing and unclear; there is little to no agreement on the true sustainability drivers (Billio et al., 2021), which are moreover plagued by missing data at the company level. As a consequence, there is no such thing as a best practice, even for large-cap companies, for which there are long time series available and much more information has been released over the last decade, especially if compared to SMEs, and this holds true when dealing with missing data as well. As (Kotsantonis and Serafeim, 2019) argue, the disclosure of ESG issues tends to be much more limited in smaller companies, making the data imputation even more important for SMEs. However, we emphasize that, when it comes to the availability of statistical techniques, lack of long time series and poor data quality make multiple imputation techniques difficult to apply for hand-collected data from SMEs.

Finally, as for the management of missing data, there is one further point to stress. When attributing a score to a firm, it is crucial to determine whether a company should receive a penalty for not disclosing the data or not. For instance, (Sahin et al., 2022) introduce a fourth pillar, called the 'missing' (M) pillar, where the higher the number of missing data points, the higher its overall ESGM score. The authors argue that this approach is designed to soften the impact of ESG exclusion strategies, which might remove by construction companies with potentially high scores, also because unavailability does not imply necessarily unwillingness or inability to release data. The general goal of such an approach is rather to encourage companies to release ESG data.

Since we deal with a sizeable amount of undisclosed data, we decide to attribute the worst sector value to companies that do not release data, and we consider two ways of imputing missing data for a given criterion. As we have previously argued, small firms might be more reluctant to release data due to direct and indirect costs. Furthermore, the underlying idea is born from the fact that the missing information could actually be a negative signal about the firm, and imputing an average score might be a far too positive assumption, so it might be safer to assume that it is doing poorly (Lindsey et al., 2022), given both the characteristics of the data and the size of the firms involved. Therefore, in the first case, we impute the firm with the worst sector value, whereas in the second case, if there are no available data points in a given sector, we impute the worst global value. Although we recognize that this choice might be an excessively harsh imputing procedure for small businesses, this assumption is mainly justified by the necessity of ensuring that scores are not magnified by loose imputation rules. Since other statistical approaches, such as predictive mean matching or multiple linear regression, are not feasible in this context due to lack of data, we find a sweet spot between mean and multiple imputations for handling this problem by applying a prudential rule, although it is well known that mean (worst-value) replacement tends to reduce (magnify) the volatility of variables, and as a result, variances and covariances are also under(over)-estimated, resulting in biased estimators, especially if data are not missing completely at random (MCAR) (Enders, 2022).

Nonetheless, although we make a relatively safe and prudential assumption, we cannot ignore that our sample might not accurately represent the behaviour of the *entire* universe of SMEs. Hence, under strict and specific modeling assumptions, it might not be convenient to disclose sustainability information voluntarily; in Section 5.3, we also aim to quantify such an effect. Therefore, it is not completely unreasonable to assume that if the model design is publicly available, then poorly performing firms, i.e. firms whose alternatives $g_{ij,\theta} \notin \mathcal{G}$, where θ and $\mathcal{G}$ denote respectively the sector and the set of $i = 1, \dots, n$ companies releasing data for the j-th alternative, are such that $g_{ij} < min(g_{ij,\theta}) \in \mathcal{G}$, might not be encouraged to report such information, resulting in a self-selection bias. From a normative point of view, this provides a case against attributing an additional penalty for missing observations, and this is also the line of reasoning of the ESGM approach (Sahin et al., 2022). In spite of that, we stress that our assumption is relatively strong and might not fully capture the behaviour of poorly performing companies taking advantage of a fully transparent rating methodology.

Finally, note that we do not control for system-wide effects. Lack of disclosure is also essential when assessing the global impact of the scoring procedure. Our concern is, in particular, related to the left tail of the distribution of scores, which is far from being stable due to the unwillingness or inability of poorly performing companies to disclose ESG data. Such behaviour makes our estimates either upward or downward biased. On the one hand, we stress the presence of an upward bias since the *true* distribution of missing data might be actually worse than estimated and more skewed to the left. On the other hand, the fact that some companies on the left tail of the distribution do not provide additional information has a negative impact on both the computed inflows and outflows, which ultimately affect the final estimated scores.

3.2 Sustainability indicators

Determining proper sustainability indicators that are informed by different voluntary reporting standards and frameworks across sectors and countries is a complex task, which can be basically boiled down to (1) assessing the double materiality of each topic and (2) slimming down the list of candidate topics in order to establish a level playing field for all companies that voluntarily disclose and report their corporate social responsibility policy.

Although both topics are considered of paramount importance in the literature (Khan et al., 2016), due to limited willingness to disclose CSR information caused by a poor cost-benefit ratio of CSR reporting for microcap companies, a thorough assessment of relevant criteria becomes quite complicated (Gjergji et al., 2021).

Table 1: Sample size of companies disclosing ESG information by country and sector. Note that EU countries without relevant data are not included in the table; in parenthesis, the overall number of listed SMEs in the EU is reported. The clustering of firms by sector is the result of a NACE-to-GICS mapping performed by the authors.

Country \ Sector	Industrials	Consumer discretionary	Consumer staples	Health Care	Financials	Information Technology	Communication services	Utilities	Real Estate	**Total**
Austria				1						1 (4)
Belgium	1								1	2 (19)
Denmark	4	1			5	2			1	13 (59)
Estonia									1	1 (8)
Finland	2			1	1	1				5 (37)
France	5			5	1	1		1	1	14 (191)
Germany		1	1			1	1			4 (120)
Hungary					1			1		2 (15)
Italy	3	5		3	2	8	3	1	1	26 (226)
Latvia		1								1 (5)
Netherlands			2							2 (18)
Poland									3	3 (211)
Romania	1									1 (97)
Spain								2	1	3 (46)
Sweden	10	2		3	2	3			6	26 (220)
Total	26	10	3	13	12	16	5	5	15	**104 (1,337)**

Therefore, a non-discretionary assessment of the relevant criteria is necessary: a thorough screening of the sustainability reports from an initial sample of 1,337 listed European SMEs has been performed. The choice of criteria has neither been driven by the sample characteristics nor by sector or country-specific factors, but it has been rather influenced by an overall evaluation of all the reports according to (Global Reporting Initiative, 2022) standards. After detecting a pool of possible candidates of criteria, we subsequently filter them by removing all the criteria characterized by a boolean value (e.g. having a whistle-blowing system in place), since after a comprehensive screening of all the sustainability reports, they were found to provide little added value to the analysis.

The description and the purpose of criteria is reported in Table 2, where the environmental, social and governance topics are respectively denoted as E_{ij}, S_{ij} and G_{ij}, and i, j identify respectively the criteria and the alternatives. The main idea is to capture the behaviour of the company across three dimensions in a comprehensive and parsimonious way. Hence, we identify 12 topics that are generally considered material also for large cap companies, so as to encourage convergence towards the reporting standard of larger firms.

We end up with a unique hand-collected dataset with 811 observations (104 firms observed across 12 different criteria, with around 35% missing data) for a cross-sectional study. The screening process furthermore required, whenever possible, adjustments of the reported number so as to guarantee homogeneous comparisons among firms. If the disclosed data of a firm cannot be harmonized in accordance with the provided definitions, then the value is reported as missing.

Furthermore, companies are included in the new sample only if minimal reporting standards are met, i.e. more precisely if accurate information for at least seven criteria out of twelve is supplied. In this way, we aim to minimize potential inconsistencies and subjectivity of results. All the criteria are constructed in accordance with the requirements of (Global Reporting Initiative, 2022) reporting standard.

A summary of the outcome of the screening process is reported in Table 1. Moreover, in Figure 1 we show that distributions of criteria are mostly bimodal: note that the imputation procedure can significantly affect the shape of the distribution of specific criteria.

Table 2: List of criteria. The cardinality, the goal and a concise description are respectively reported, along with the identification of each criterion, for the alternative j.

Criterion	Cardinality (#)	Goal	Description and purpose
E_{1j}: Carbon intensity	70	Min	Carbon intensity captures both direct emissions from owned or controlled sources (scope 1) and indirect emissions from the generation of purchased electricity (scope 2). It corresponds to the ratio of Greenhouse Gases (tCO_2e) emissions to the firms' revenues in €. Scope 3 emissions (i.e. emissions indirectly generated by the company's value chain) are not considered due to lack of reliable estimates, nonetheless this measure allows us to gauge approximately the firm's exposure to climate risk.
E_{2j}: Waste generation intensity	44	Min	The waste generation intensity is defined as the ratio of hazardous and non-hazardous waste generation measured in *tonn.* and the firm's revenues in €. For real estate companies, both the head offices and the investment properties (indirect) waste generation is considered. Its relevance is related to climate risk exposure.
E_{3j}: Non-renewable electricity consumption (%)	95	Min	The criterion is computed as the ratio of non-renewable electricity to total electricity consumption (in KWh), with the aim of capturing the firm's commitment to an environmentally sustainable supply chain.
E_{4j}: Water consumption intensity	48	Min	The water generation intensity is defined as the ratio of water consumption in m^3 to firm's revenues in €. For real estate companies, the investment properties water consumption is also considered. In this case, the firm's exposure to climate risk is assessed.
S_{1j}: Average training hours	40	Max	The criterion corresponds to the ratio between overall training hours (vocational training, instruction and training or education pursued externally) and the overall headcount (middle and top management included) at time t. In this way, commitment to high quality working conditions is captured.
S_{2j}: Job creation (%)	94	Max	The criterion corresponds to the ratio of the difference between new hires and terminations at time $t+1$ to the overall headcount at time t (interns included). The ratio aims at gauging the commitment of the company at achieving a sustained economic growth.
S_{3j}: Management diversity by gender (%)	83	Max	The criterion is defined as the ratio of female managers to the total number of managers minus 50% to ensure gender parity among managers. Both middle and top management are included.
S_{4j}: Gender pay gap	20	Min	The gender pay gap is defined as the absolute difference between average gross annual wage of male and female employees, without controlling for seniority, so as to evaluate the firm's commitment to gender parity from a different standpoint.
G_{1j}: Board diversity by gender (%)	104	Max	The criterion is defined as the ratio of female directors to the total number of directors minus 50% to ensure gender parity among directors.
G_{2j}: Economic value generation and distribution (%)	72	Max	The criterion is defined as the ratio of economic value distributed by the company to its stakeholders such as suppliers, employees, lenders, public administration and shareholders to the economic value generated, i.e. the annual turnover. This value represents the wealth produced by the firm and its impact on key stakeholders.
G_{3j}: Board independence (%)	104	Max	The criterion is defined as the ratio of independent directors to the overall number of directors. Such a criterion is aimed at providing a rough assessment of the quality of corporate governance.
G_{4j}: CEO pay ratio (%)	37	Min	The criterion corresponds to the ratio of the annual remuneration paid to the CEO to the average annual remuneration of all employees. The goal is to measure the commitment of the company at mitigating income inequality at firm and society levels.

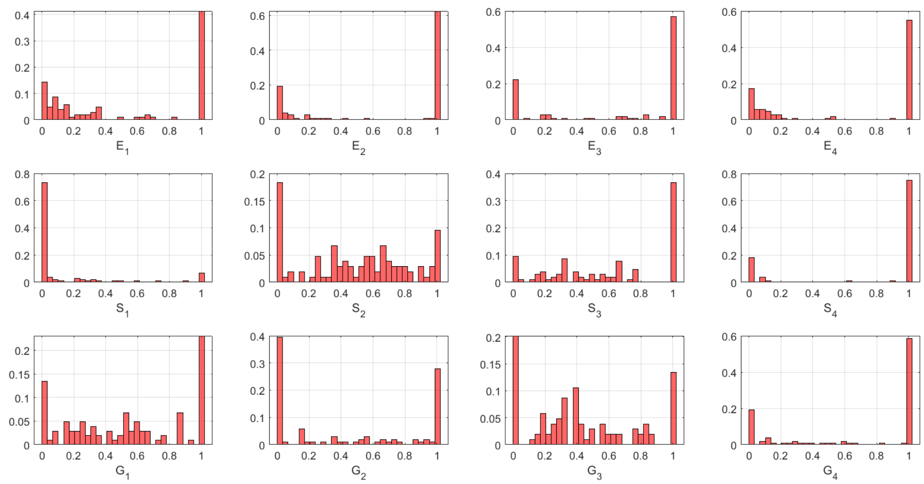

Figure 1: Criteria distribution estimate via histograms w.r.t. normalized and imputed datasets. The impact of the single imputation method can be observed either on the right or the left tail of the distribution, respectively, depending on whether the criteria is minimized (E_1, E_2, E_3, E_4, S_3, S_4, G_1, G_4) or maximized (S_1, S_2, G_2, G_3).

As far as the disclosure process is concerned, please note that the breakdown of the sample reported in Table 1 by sector and country is characterized by significant concentration, with two countries making up exactly the 50% of the overall sample size and four countries, that is to say, Sweden, Denmark, Italy and France, represent more than 75% of the sample size. Despite considering a small sample, our results are aligned with findings in the literature and could be explained in various ways. The standards and the intensity of ESG reporting differ significantly across countries, industries, and firms. Moreover, it has been shown that peer effects and the stakeholders' commitment to ESG targets and CSR strategies carried out by competitors in a given ecosystem are known for impacting on the quality of ESG reporting (Torelli et al., 2020).

Altogether, we document that companies are not particularly willing to disclose information related to their employees' pay, in particular when taking into account the gap between female and male employees, as well as with their CEOs. Furthermore, on the one hand, we find that the number of missing values is approximately stable across countries, whereas we find that the CEO pay ratio can be easily recovered from mandatory disclosure for Swedish companies, so the replaced ratios are substantially impacted by a strong asymmetry in reporting standards. In other words, we are basically assuming that the ratios for most non-Swedish companies are roughly equal to the Swedish's highest ratio between CEO and employees' pay for a given sector, and this might potentially hide even worse values, explaining, as a result, the lack of disclosure for that specific topic.

As for the remaining topics, we find that a relatively high percentage of companies disclose sustainability-related information, although some criteria, such as the number of independent directors, are also affected by mandatory requirements. Finally, we find a relatively high number of missing values for other criteria, such as the number of training hours per employee, which should not be considered particularly concerning since a full screening process simply revealed that the topic in most cases was not considered material at the time of releasing: for instance, w.r.t. to the training hours, more than 50% of the available data is referred to Italian companies: we conclude that the numbers are also affected by peer-effects, country, and ecosystem-specific reporting standards. In addition, aside from mandatory requirements, we find that companies tend to limit disclosure to a restricted scope in relation to governance topics. The resulting difficulty in extracting drivers for assessing the governance performance could also be related to the negative relationship between firm performance and corporate governance quality (Guney et al., 2020) found in the literature, especially with respect to board diversity in SMEs (Shehata et al., 2017).

4 A multicriteria decision model for ESG assessment

In light of the lack of quantitative approaches for modeling ESG profiles for SMEs, in this section, we propose to cope with such a problem by using an MCDA approach (Zionts, 1979). MCDA methods allow us to deal with unstructured problems in a flexible way by assessing the impact of multiple conflicting criteria in decision-making. More precisely, in what follows, we resort to the MURAME model (Goletsis et al., 2003), which allows us to make a very limited number of assumptions with regard to the involved parameters and missing data. At the same time, MCDA approaches are more accurate and have better discriminatory power than composite performance indicators based on standard scorecards (Siddiqi, 2017).

As (Keeney and Raiffa, 1993) state, the theory of decision analysis is designed to assist the DM in making a decision among a set of competing alternatives when conflicting objectives are present. Among the many available MCDA methods for dealing with multi-criteria decision problems, we choose a baseline model, the multicriteria ranking method (MURAME) (Goletsis et al., 2003), for various reasons. Compared to other MCDA models, the MURAME belongs to a family of multicriteria models (Brans and Vincke, 1985) for which only a limited number of assumptions is necessary. Other popular methods, such as the MAUT (Keeney and Raiffa, 1993), require the DM to express his utility function. As we discuss below, we do not want to rely on expert judgment in order to reduce subjectivity in the decision-making problem as much as possible. The degree of subjectivity in the parameter setting is also kept to a minimum, as we document in Section 5.1. as we want the results to be interpretable and transparent. The flexibility of our approach also allows for a natural extension to more refined scoring procedures, including further alternatives, as more detailed sustainability reports are expected to appear in the next few years. Hence, in light of the insufficient or unreliable track record of SMEs w.r.t. sustainability reporting, the selected approach seems particularly suitable for the discussed problem.

In what follows, we illustrate the methodology used in this work and first proposed by (Goletsis et al., 2003), a multicriteria outranking methodology (MURAME) which merges two deep-rooted multi-criteria decision-aiding methodologies, namely ELECTREE III (Figueira et al., 2016) and PROMETHEE II (Brans and Vincke, 1985). Among others, it has been applied to energy projects evaluation (Goletsis et al., 2003) and credit risk assessment (Corazza et al., 2015, 2016).

The presentation follows closely the discussion of the methodology in (Goletsis et al., 2003). In Section 4.1, we illustrate the methodology, which requires the definition of a preference structure and an outranking relation to recover the final ranking of the alternatives. Finally, in Section 4.2, we relate the characteristics of the data to the model.

4.1 A multicriteria ranking method (MURAME)

Let us consider first the preference structure of a decision-maker. Given a set of m alternatives $\mathcal{A} = \{a_1, \ldots, a_i, \ldots, a_m\}$ and a set of n criteria $\mathcal{C} = \{c_1, \ldots, c_j, \ldots, c_n\}$, the alternatives are assessed according to a decision matrix $\mathbf{G}_{m \times n}$, whose elements g_{ij} contain the score for the alternative i and the criterion j. In order to reflect the actual DM's preference structure, weak preferences, as well as more realistic and fuzzy preference relations, can be adopted.

The DM's uncertain preferences are taken into account by means of the introduction for any given criterion c_j, of an indifference threshold, q_j, and a preference threshold, p_j, so as to allow for three different preference zones. Namely, the DM can express a sure preference, uncertain preference, or indifference among two given alternatives. In any case, it holds that $q_j \leq p_j$.

In the second step, an outranking relation is defined by constructing, for each $a_i, a_k \in \mathcal{A}$

and with $i \neq k$, an outranking relation, according to the above-mentioned preference structure. The outranking relation is derived by computing specific indexes, namely the concordance and discordance indexes, which provide quantitative measures of the degree of dominance of an alternative over another.

Let us begin by defining, for each pair of alternatives, the local concordance $C_j(a_i, a_k)$ and discordance $D_j(a_i, a_k)$ indexes in Equations (2)-(3), as follows:

$$C_j(a_i, a_k) = \begin{cases} 1 & g_{k,j} - g_{i,j} \leq q_j \\ 0 & g_{k,j} - g_{i,j} \geq p_j \\ \dfrac{g_{i,j} - g_{k,j} + p_j}{p_j - q_j} & \text{otherwise} \end{cases} \tag{2}$$

$$D_j(a_i, a_k) = \begin{cases} 0 & g_{k,j} - g_{i,j} \leq p_j \\ 1 & g_{k,j} - g_{i,j} \geq v_j \\ \dfrac{g_{k,j} - g_{i,j} - p_j}{v_j - p_j} & \text{otherwise} \end{cases} \tag{3}$$

If $g_{k,j} \geq g_{i,j} + p_j$, then the DM expresses a strict preference for the alternative a_k over the alternative a_i. Therefore, if the local concordance index reaches its minimum value for any given pair of alternatives, then a_i is dominated by a_k. Conversely, if $g_{k,j} \leq g_{i,j} + q_j$, then a_k is not preferred to a_i, resulting in the local concordance index of pair a_i, a_k reaching its maximum value. In the MURAME design, a degree of fuzziness in the DM's preferences is introduced by defining a region in which he holds a weak preference w.r.t. an alternative over another, hence the local concordance index $C_j \in (0, 1)$.

To further assess the deviation from the hypothesis that a_i dominates a_k, according to the criterion c_j, a discordance index $D_j(a_i, a_k)$ is correspondingly introduced so as to quantify the extent to which the dominance assumption is not satisfied. A third threshold v_j such that $v_j \geq p_j$, also known as veto threshold, is retained in order to reject the hypothesis that the alternative a_i is at least as good as alternative a_k: if so, the discordance index attains its maximum value.

We proceed now by setting up the global concordance index in Equation (4) through the aggregation of the local concordance indexes, as follows:

$$C(a_i, a_k) = \sum_{j=1}^{n} w_j C_j(a_i, a_k), \tag{4}$$

where w_j represents the normalized non-negative weight associated with a given criterion j.

An outranking index $O(a_i, a_k)$ is constructed in Equation (5) by putting together information from both discordance and global concordance indexes, indicating to what extent the alternative a_i outranks a_k, for all j:

$$O(a_i, a_k) = \begin{cases} C(a_i, a_k) & D_j(a_i, a_k) \leq C(a_i, a_k) \quad \forall j \\ C(a_i, a_k) \displaystyle\prod_{j \in \mathcal{J}} \frac{1 - D_j(a_i, a_k)}{1 - C(a_i, a_k)} & \text{otherwise,} \end{cases} \tag{5}$$

where $\mathcal{J} \subseteq \{1, \ldots, n\}$ is the subset of criteria such that $D_j(a_i, a_k) > C(a_i, a_k)$. The outranking index is equivalent to the global concordance $C(a_i, a_k)$ index unless the performance of an alternative, in relation to at least one criterion, is poor to the extent that it poses a veto on the global outranking relation, determining a subsequent decrease in the outranking index. If there exists even a single criterion for which the discordance index $D_j(a_i, a_k)$ hits the maximum value, i.e. 1, for a given criterion j, the outranking index (5) is equal to zero.

Finally, we put together the computations performed in the two previous steps to recover a final ranking of alternatives. This can be attained by computing a difference which is also known as net flow, between in and out-flows, for each alternative a_i. The outflow $\varphi^+(a_i)$ quantifies the comparative strength of the alternative a_i when compared to the remaining alternatives. Conversely, the inflow, denoted as $\varphi^-(a_i)$, captures the relative weakness of alternative a_i by evaluating the comparative strength of all other alternatives in relation to it:

$$\varphi(a_i) = \varphi^+(a_i) - \varphi^-(a_i) = \sum_{k \neq i} O(a_i, a_k) - \sum_{k \neq i} O(a_k, a_i). \tag{6}$$

Hence, the alternatives can be ranked according to the net flow φ and normalized so as to attach a score $\mathcal{S} \in [0, 100]$ to each alternative.

4.2 Relating the model to the data

The MURAME is a parsimonious model requiring a limited number of assumptions w.r.t. to the DM's preferences. However, the characteristics of the data and the imputation method can still have a notable impact on intermediate computations, hence a better understanding of the relationship between the criteria and the final net flows is of paramount importance. As for the sensitivity to the imputation method and the related implications, we refer the reader to Section 5.2.2. In what follows, we expand on the relationship between the data and the proposed model.

We refer to Table 2 for a general description of criteria, whereas in this section we comment on the formal construction of each criterion in Table 3, along with the normalization criterion. For three out of four environmental criteria, we normalize by company revenues in million euros. The third criterion, E_{3j} is equal to the percentage of non-renewable electricity consumption. As for the social criteria, we divide by the number of employees (see e.g. criteria S_{1j} and S_{2j}) or we consider the absolute value of a distance to assess gender parity. The definition of governance normalization criteria is also straightforward, since we consider the absolute distance from gender equality, as well as revenues and the number

Table 3: Measurement of criteria, where $i = 1, \ldots, 4$ for each criterion and j denotes alternatives.

Criterion	Measurement
E_{1j}: Carbon intensity	$\mathcal{CI}_j = \frac{\mathcal{CE}_j}{\mathcal{R}_j}$
E_{2j}: Waste generation intensity	$\mathcal{WGI}_j = \frac{\mathcal{WG}_j}{\mathcal{R}_j}$
E_{3j}: Renewable electricity consumption (%)	$\mathcal{RS}_j = \frac{\mathcal{RE}_j}{\mathcal{E}_j}$
E_{4j}: Water consumption intensity	$\mathcal{WCI}_j = \frac{\mathcal{WC}_j}{\mathcal{R}_j}$
S_{1j}: Average training hours	$\overline{\mathcal{T}}_j = \frac{\mathcal{T}_j}{\mathcal{EM}_j}$
S_{2j}: Job creation (%)	$\mathcal{JC}_j = \frac{\mathcal{EM}_{t,j} - \mathcal{EM}_{t-1,j}}{\mathcal{EM}_{t-1,j}}$
S_{3j}: Management diversity by gender (%)	$\mathcal{MD}_j = \lvert\frac{\mathcal{FM}_j}{M_j} - 0.5\rvert$
S_{4j}: Gender pay gap	$\mathcal{GG}_j = \lvert\mathcal{MR}_j - \mathcal{FR}_j\rvert$
G_{1j}: Board diversity by gender (%)	$\mathcal{BD}_j = \lvert\frac{\mathcal{FM}_j}{\mathcal{M}_j} - 0.5\rvert$
G_{2j}: Economic value generation and distribution (%)	$\mathcal{EVD}_j = \frac{\mathcal{RD}_j}{\mathcal{R}_j}$
G_{3j}: Board independence (%)	$\mathcal{BI}_j = \frac{\mathcal{IBM}_j}{\mathcal{BM}_j}$
G_{4j}: CEO pay ratio (%)	$\mathcal{R}_{\mathcal{CEO}/\mathcal{EM},j} = \frac{\mathcal{R}_{\mathcal{CEO},j}}{\mathcal{R}_{\mathcal{EM},j}}$

of board members. The average salary is used as a standardizing criterion to compute the CEO pay ratio.

Descriptive statistics of criteria are reported in Table 4. As for the distribution of criteria, we find regular evidence of highly skewed and and fat-tailed distributions. This is also reflected by the fact that the distributions are often bimodal. Results from standard Jarque-Bera tests generally point to a significant departure from normality assumptions.

In Table 5, we report the unconditional correlations between criteria, with the associated p-values, where the test statistics $t \sim \mathcal{T}(\nu - 2)$ with $\nu - 2$ denoting the degrees of freedom. Note that in most cases, we cannot reject the null hypothesis that $\rho = 0$, with a few exceptions for variables belonging to the same pillar or variables that are related to each other by construction (e.g. carbon emissions may include a portion of waste generation).

Table 4: Descriptive statistics of data. Significance codes: *** at the 99% level; ** at the 95% level.

Criterion	Mean	St.dev.	Skewness	Kurtosis	Normality test
E_{1j}	22.1022	24.9927	1.1360	2.9254	22.3921***
E_{2j}	37.3192	53.3831	1.2419	3.0198	26.7346***
E_{3j}	0.6684	0.4315	-0.7025	1.6459	16.4996***
E_{4j}	1156.5	2816.9	3.7907	16.9799	1096.1***
S_{1j}	12.247	16.4705	3.1248	13.7082	666.12***
S_{2j}	0.0918	0.2073	1.5729	12.0204	395.48***
S_{3j}	0.2980	0.1762	-0.1172	1.5688	9.1138**
S_{4j}	0.3966	0.2575	0.3071	1.8645	7.222**
G_{1j}	0.2681	0.1417	0.2082	2.1536	3.8561**
G_{2j}	0.8660	0.1333	-0.7915	2.7288	11.1783***
G_{3j}	0.4148	0.3187	0.4543	2.1910	6.4143**
G_{4j}	8.1677	4.4245	0.3703	1.5970	10.9064***

As far as the dependence between criteria is concerned, although multicriteria decision aiding does not suffer from multi-collinearity issues, we report an ex-post assessment of Pearson correlations since:

- A large number of variables typically undermines the interpretability of the model, especially if highly correlated variables are included in the model;
- Highly correlated variables do not add particular value when performing pairwise comparisons between criteria, and such behaviour is generally detrimental to an in-depth assessment of the key drivers of the model.

Nonetheless, the MURAME, just as other MCDA models, can flexibly accommodate a large number of criteria and alternatives, regardless of the correlations among variables

(Corazza et al., 2016).

Table 5: Unconditional correlations between criteria, with p-values reported in parenthesis.

	E_1	E_2	E_3	E_4	S_1	S_2	S_3	S_4	G_1	G_2	G_3	G_4
E_1	1.000000	0.473940 (0.0053)	0.021596 (0.9395)	0.325944 (0.0509)	0.128877 (0.6134)	0.068958 (0.9812)	−0.064485 (0.3053)	−0.225304 (0.0461)	0.003578 (0.8257)	−0.121592 (0.2854)	−0.102407 (0.3891)	0.077008 (0.6077)
E_2		1.000000	−0.032055 (0.7923)	0.295190 (0.0690)	−0.001474 (0.9816)	0.188957 (0.5311)	−0.021804 (0.4285)	−0.308871 (0.0144)	0.156091 (0.6303)	−0.038047 (0.5717)	−0.141679 (0.3280)	−0.035813 (0.3470)
E_3			1.000000	0.154220 (0.5486)	−0.051077 (0.6750)	−0.212740 (0.0856)	0.064861 (0.8843)	0.006408 (0.8506)	0.167668 (0.4777)	−0.094965 (0.3789)	−0.180057 (0.2756)	−0.019551 (0.6705)
E_4				1.000000	−0.055421 (0.6928)	0.018858 (0.7075)	−0.026512 (0.364)	−0.296369 (0.0258)	−0.007542 (0.7701)	−0.139270 (0.2881)	0.047698 (0.9378)	−0.081964 (0.1870)
S_1					1.000000	0.077596 (0.8465)	−0.036117 (0.6460)	−0.093146 (0.5400)	0.065786 (0.8682)	−0.072337 (0.5737)	−0.159572 (0.2848)	0.073376 (0.9282)
S_2						1.000000	0.146203 (0.7374)	0.006118 0.7934	0.118434 (0.8128)	0.143164 (0.5665)	−0.041570 (0.6501)	0.129611 (0.7904)
S_3							1.000000	0.256714 (0.1481)	0.188442 (0.4809)	−0.063934 (0.5325)	−0.140927 (0.3041)	0.240855 (0.2013)
S_4								1.000000	−0.089472 (0.5252)	−0.069366 (0.8031)	0.061944 (0.7823)	0.358087 (0.0497)
G_1									1.000000	0.022789 (0.8357)	−0.166676 (0.1525)	0.098757 (0.9538)
G_2										1.000000	0.082461 (0.5886)	−0.004098 (0.6953)
G_3											1.000000	0.113700 (0.3583)
G_4												1.000000

5 Application to the ESG scoring problem of SMEs

In what follows, we apply the scoring procedure described in section 4 to the considered dataset of European SMEs. Before moving on to discussing the findings and the policy implication of the present study, let us point out briefly the main assumptions of our approach:

- A set of (Global Reporting Initiative, 2022)-based indicators is adopted in order to objectively assess the firms' sustainability performance and to favor alignment to international reporting standards;
- Each aspect of ESG profiles assessment is assumed to be equally important after sector-specific normalization;
- The firms' scores are recovered according to an outranking approach, where the sustainability performance is directly related to a pre-specified preference structure

of the DM;

- Due to the peculiar cost-benefit structure of voluntary ESG reporting for SMEs, a prudential heuristic is adopted to replace missing values.

Let us now illustrate the structure of the ensuing application. For the sake of readability, in this section, we focus on the top ten and worst ten performers, and we set four main goals:

- First, we break down the overall scores by pillar, and we aim to highlight the key drivers of performance of top and worst performers in Section 5.1;
- Second, the scores of Real Estate firms are fully broken down. The sector is chosen in light of a low number of missing values and due to the large volatility of scores observed across firms, for all the criteria, in Section 5.1;
- Then, we perform a sensitivity analysis in Section 5.2, by which we aim to support our claims about the robustness of a neutral and objective parameter setting;
- Finally, we make an assessment of policy implications in Section 5.3, in which we also aim to quantify the benefits or the disadvantages of disclosing ESG information under the prudential framework discussed in Section 3.1.

5.1 Scoring procedure and breakdown by pillar

In this section, we discuss the performance of firms across different sustainability dimensions, hence multiple tests are run to capture both the overall standings and the performance within each pillar.

As for the parameter settings, in the same spirit of the discussion in Section 3.1, we go for a neutral approach and keep subjectivity at a minimum. Therefore, equal weights are applied to each criterion, and the thresholds q_i, p_i, and v_i are set respectively equal to the first, the third, and the fourth quintile of the distribution of criterion i.

In Table 6, we present the results with regard to the top ten and worst ten performers. We use a palette of ten colors to identify exactly ten ranges of normalized scores $\mathcal{S} \in [0, 100]$, from $[0, 10)$ to $[90, 100]$. Furthermore, note that by construction, there is not a linear relationship between aggregate ESG scores $\varphi_{net,ESG}$ and the scores for each pillar $\varphi_{net,i}$, with $i \in \{E, S, G\}$ (see Equations (4)-(5)). Indeed, the scores of the E, S, G columns are obtained by running the model three times separately for each pillar i, where we set the weight for pillar i (i.e. a subset of criteria belonging to i). The preference of the decision maker w.r.t. two or more alternatives can be measured and quantified, although only within each pillar, hence the column ESG in Table 6 should not be interpreted as the weighted average of the other three columns. Indeed, by attributing a weight only to a single pillar i, the sustainability performance of firms in columns E, S, G should be interpreted as if the DM attached importance to only one of the three profiles.

As for the results in Table 6, note that the first firm systematically outperforms the others by a wide margin, also thanks to a strong performance across all dimensions, since it is ranked in the top five also within single sustainability dimensions. Among the leading firms, the utilities and communication services sectors stand out. As for the former, a preliminary analysis of all the collected sustainability reports shows that renewable energy companies are more prone to release ESG data and are generally posed to outperform peers, especially from an environmental point of view. With respect to geographical and sectoral clusters, note also that all the Spanish firms included in our sample are ranked in the top ten, whereas among the laggard firms, note that Information Technology SMEs, despite releasing more information compared to firms in other sectors, tend to perform poorly on average. Altogether, due to both the imputation procedure and the outranking nature of

Table 6: Top and bottom ten firms according to the ESG scoring procedure. The overall score is reported along with a breakdown by pillar. Please note that, by construction, the overall score based on the net flow φ_{net} does not correspond to the weighted average of the score of each pillar.

Ranking	ID	Sector	Country	ESG	E	S	G
1	57	Utilities	Spain	100.00	100.00	100.00	93.79
2	91	Industrials	Sweden	77.92	100.00	75.71	77.21
3	42	Industrials	Finland	76.12	86.94	78.78	78.09
4	11	Communication Services	Italy	71.88	72.42	89.73	76.06
5	58	Utilities	Spain	67.45	80.61	78.94	74.53
6	70	Industrials	France	66.81	81.12	75.23	66.34
7	9	Communication Services	Italy	66.11	80.07	82.38	73.04
8	93	Industrials	Sweden	65.32	65.07	61.08	88.28
9	56	Communication Services	Spain	64.59	66.20	66.06	88.95
10	8	Consumer Discretionary	Italy	64.16	99.59	83.39	37.01
95	68	Information Technology	France	22.51	35.52	40.54	32.67
96	50	Industrials	Netherlands	21.21	0.87	37.29	38.17
97	13	Real Estate	Italy	20.88	0.75	16.56	37.93
98	38	Information Technology	Denmark	17.11	31.19	31.31	24.15
99	16	Financials	Italy	13.36	0.31	52.89	18.02
100	26	Health Care	Italy	10.58	0.85	39.83	20.62
101	23	Information Technology	Italy	1.93	0.04	18.61	14.20
102	75	Industrials	Romania	1.15	16.08	4.74	15.97
103	53	Consumer Discretionary	Germany	0.91	0.00	18.58	0.00
104	28	Financials	Denmark	0.00	14.81	15.77	15.44

the model, the degree of voluntary disclosure achieved by European companies reporting on sustainability issues is definitely a crucial determinant of rankings. The results in Table 6 are therefore significantly affected by lack of data, especially with respect to social and governance pillars, for which the reporting standards are found to be generally lower.

Recall that we implicitly assume that it is important for a firm to deliver comprehensive reporting and to be consistent across the three sustainability dimensions, irrespective of sectoral peculiarities, by attributing equal importance to each sustainability dimension of the firm. This explains in part why some companies that may be focused on reporting on a specific sustainability dimension, which is deemed to be more material than others, tend to underperform: for instance, note that the best-performing company from a governance point of view does not appear in the top ten. The relationship between the pillar scores and the overall ESG rating is also highlighted in Figure 2: a broad range of possible combinations of scores emerges, as well as the pronounced outperformance (underperformance) of best (worst)-performing companies.

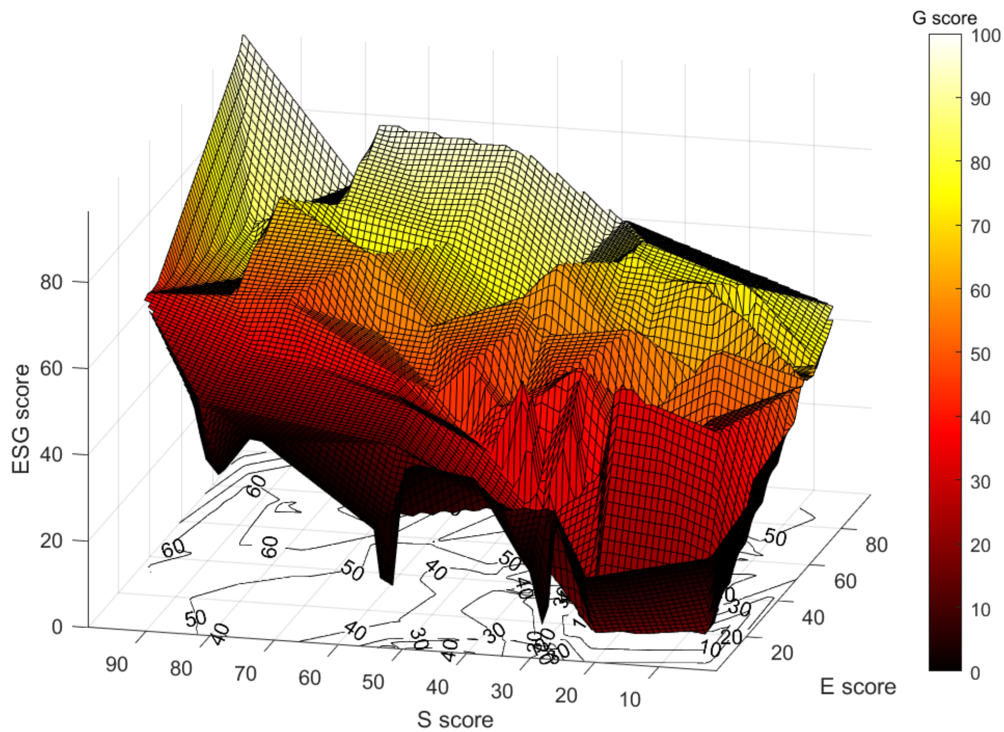

Figure 2: Relationship between ESG, E, S and G scores. The contour lines represent the values of the ESG aggregate score, while we report respectively on the x-axis and the y-axis the E and S scores. The colors are used to characterize the value of the G score.

Given the large number of real estate firms reporting ESG data and the significant heterogeneity in reported rankings, we propose to briefly focus on such firms to better understand the drivers of their ESG performance. Since such a sector comprises very different businesses and is actually found to be more sensitive to different parameter settings, it is worth proposing a focused analysis. In addition, some important distinguishing traits of REITs with regards to ESG performance have been recently documented, involving, in particular, the (positive) relationship between sustainability achievements and occupancy rates, property prices, and debt financing (Feng and Wu, 2021). With respect to our results, in Table 7 a breakdown of the scores is proposed. High variability and lack of consistency emerge across the three sustainability dimensions. Furthermore, poor or insufficient track records with respect to specific pillars emerge for a few companies, which receive scores close to zero. Therefore, also in this case, we observe much lower scores for both the social and the governance pillars, mainly due to a high number of missing observations, whereas the worst-ranked firm does not actually report on environmental issues.

Table 7: A focus on real estate firms. The overall score is reported along with a breakdown by pillar. Please note that, by construction, the overall score based on the net flow φ_{net} does not correspond to the weighted average of the score of each pillar.

Ranking	ID	Sector	Country	ESG	E	S	G
16	51	Real Estate	Belgium	59.68	73.17	76.38	28.45
31	88	Real Estate	Sweden	52.50	63.20	44.80	61.11
42	76	Real Estate	Poland	49.26	58.31	64.76	27.69
46	95	Real Estate	Sweden	48.11	63.20	3.89	54.07
59	35	Real Estate	Denmark	43.80	52.84	1.00	50.33
64	101	Real Estate	Sweden	41.26	56.37	2.61	47.46
65	96	Real Estate	Swedeny	41.15	56.37	13.17	46.90
67	47	Real Estate	Estonia	40.70	29.18	68.17	52.69
69	71	Real Estate	France	40.06	43.61	69.65	34.52
79	99	Real Estate	Sweden	34.71	49.31	32.61	36.06
86	78	Real Estate	Poland	30.13	36.34	39.83	3.18
88	100	Real Estate	Sweden	29.16	39.47	13.12	31.71
97	13	Real Estate	Italy	20.88	0.75	16.56	37.93

Our results are definitely grounded in a specific and crucial assumption w.r.t. the materiality of topics. Since we do not actually assess the materiality of ESG profiles, we assume that they are equally important: it follows that more emphasis is put on the ability of a company to perform well across all the dimensions. For instance, a few companies doing particularly well from a governance standpoint do not appear among the top ten companies, indeed, due to poor performance in the remaining dimensions.

Nonetheless, as we discuss in Section 5.2, the performance of a few companies is found to be relatively insensitive to different parameterizations. Also, for the remaining companies, by letting the materiality of ESG profiles vary randomly, we do not observe extreme variations in the final ordering of firms. A materiality assessment is definitely important to establish a connection between the importance of each topic, according to its underlying economic and financial impact on the business of a company, and the influence of the firm on the ecosystem through its sustainability policies. However, we show that a good degree of robustness across a relatively wide range of settings can be achieved. As a consequence, we avoid introducing subjectivity in our results, as we further discuss in the next section.

5.2 Sensitivity analysis

In what follows, we aim to document the impact of variations across the parameter space, how the underlying model assumptions affect both the robustness of the preference ordering and the scores reported in this study. In order to assess and disentangle the impact of each parameter, we perform a sensitivity analysis. First of all, recall that the model parameterization is parsimonious, i.e. it includes:

- A vector of weights $\mathbf{w} = (w_1, \ldots, w_n)$, such that $\sum_{i=1}^{n} w_i = 1$ and $w_i \geq 0 \quad \forall i$;
- A vector of indifference thresholds $\mathbf{q} = (q_1, \ldots, q_n)$;
- A vector of preference thresholds $\mathbf{p} = (p_1, \ldots, p_n)$, such that $p_i > q_i \quad \forall q_i, p_i$;
- A vector of veto thresholds $\mathbf{v} = (v_1, \ldots, v_n)$, such that $v_i > p_i > q_i \quad \forall v_i, q_i.p_i$.

Henceforth, we denote with s the number of simulations performed in the ensuing tests.

In what follows, we analyze the sensitivity of the model with reference to two sources of uncertainty: different parameter settings and imputation procedure.

5.2.1 Assessing the robustness of the model parameterization

We employ an All-(factors)-At-a-Time (AAT) approach (Pianosi et al., 2016), in order to assess the model sensitivity to parameter variations. In AAT methods, output variations are induced by altering all the input factors simultaneously. Compared to One-(factor)-At-a-Time (OAT) methods, this approach allows for modeling both the direct influence of the perturbation of a factor, as well as the joint influence due to interactions with other factors. For simplicity, we assume that all the input factors are simultaneously drawn from independent uniform random variables.

We perform $s = 10,000$ Monte Carlo simulations aimed at identifying different regions in the inputs space corresponding to particular values of the output. After testing the MURAME across different ranges of thresholds and for all the admissible values in the feasible region of weights, we report the results for reasonably broad ranges of values, as follows:

- $w_i \in [0,1]$, s.t. $\sum_i^n w_i = 1$ and $w_i \geq 0 \quad \forall i$;
- The indifference threshold q_i is tested in a range between the 15th and the 25th percentile, the preference thresholds p_i between the 55th and the 65th percentile and finally the veto threshold v_i between 75th and the 85th percentile of criterion i.

Some preliminary tests show that:

- The impact of weight settings on scores and rankings is mostly firm-specific and somewhat sector-specific. No country-specific clustering is observed. As for the stability of the scores, we find that top performers are quite robust to perturbations

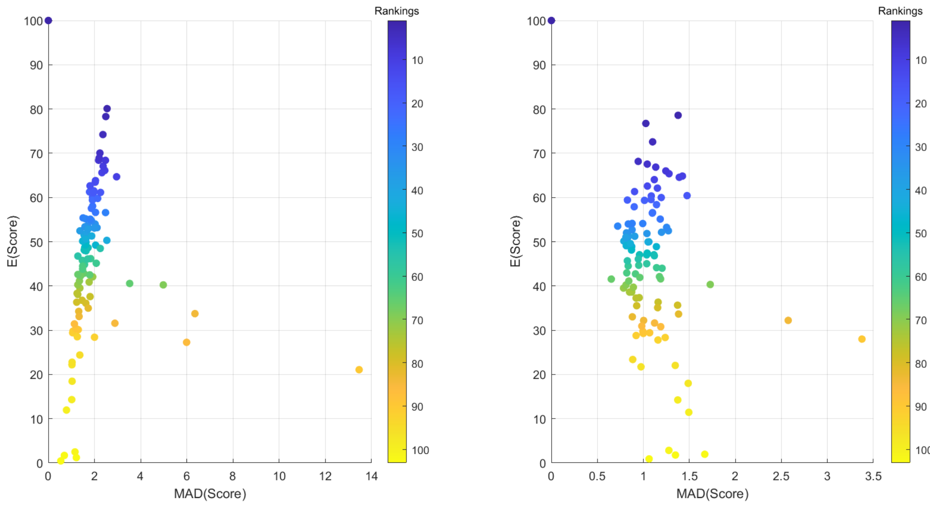

Figure 3: Sensitivity of firms' scores to thresholds settings (left panel) and across different weighting schemes (right panel) in a Mean-MAD framework. Colors in both panels are used to denote the sensitivity of rankings.

in weights, whereas we observe an increase in mean-absolute deviation (MAD) for poorly performing firms; nonetheless, the overall mean-absolute deviation (MAD) is rather low;

- With respect to the threshold settings, specific clustering effects across countries and sectors are less clear. In this case, top performers seem to be slightly more sensitive to variations in the value of the thresholds, although we observe a few outliers among worst-performing firms with a large mean-absolute deviation. The results are reported in Figures 3, 4, 5.

We also document the properties of rating assignments by assessing two basic measures of uncertainty and robustness of the assignments. Following (Doumpos and Figueira, 2019), we compute first the range of the assignments $\bar{\mathcal{R}} \in [0, 103]$ w.r.t. the attained rankings $k = 1, \ldots, m$, with $m = 104$, across a set of simulations of weights, for a given specification

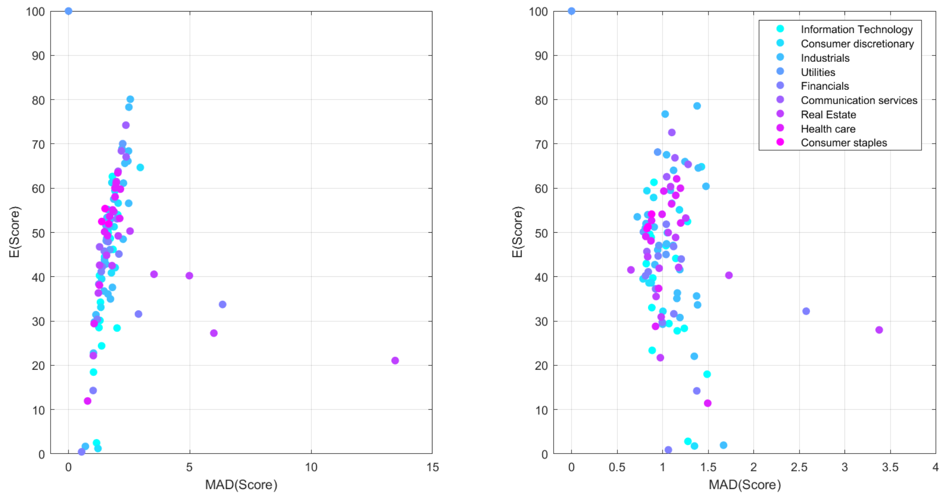

Figure 4: Sectoral clusters of firms with respect to sensitivity to thresholds settings (left panel) and weights settings (right panel) are reported, in a Mean-MAD framework. Roughly, some firms have a clustering tendency. Note that the number of observations across different sectors may vary significantly.

of thresholds, of size $s = 10,000$. It corresponds to the average range across all m firms, with $\mathcal{U}$ and $\mathcal{L}$ denoting respectively the maximum and the minimum ranking for a given company i over s simulations:

$$\bar{\mathcal{R}} = \frac{1}{n} \sum_{i=1}^{m} (\hat{\mathcal{U}}_i - \hat{\mathcal{L}}_i) \tag{7}$$

We also compute the entropy $\bar{\mathcal{E}} \in [0,1]$ of the assignments w.r.t. to the attained rankings $k = 1, \ldots, m$ across s simulations: a low entropy indicates a robust ordering across all simulations s of weights, for a given specification of thresholds; high entropy indicates instead a high level of variability in the results:

$$\bar{\mathcal{E}} = \frac{1}{m} \sum_{i=1}^{m} \Big[- \frac{1}{ln(m)} \sum_{k=1}^{m} p_{ik}^{s} ln(p_{ik}^{s}) \Big] \tag{8}$$

Table 8 summarize the results for five different settings of the thresholds, including some

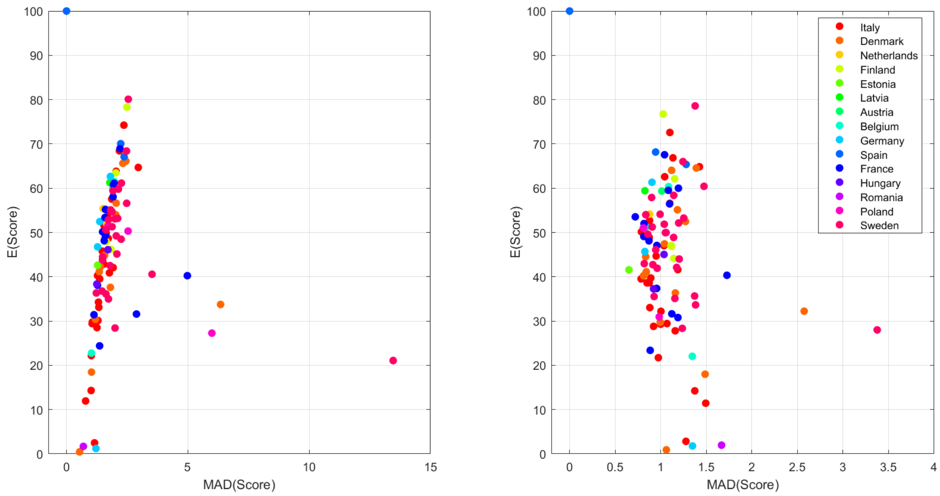

Figure 5: Geographical clusters of firms with respect to sensitivity to thresholds (left panel) and weights (right panel) are reported, in a Mean-MAD framework. No country effects emerge from data. Note that the number of observations across different countries may vary significantly.

extreme combinations of indifference, preference and veto thresholds and we document that, as long as settings are specified within reasonable ranges, a low degree of uncertainty and high robustness can be attained. In particular, the third specification seems to us a natural choice for being relatively stable, neutral and for introducing also a degree of fuzziness in the DM's preferences, by setting $q > 0$.

To assess the influence of both weights and thresholds, we also plot a stacked area chart in Figure 6, where each color denotes the overall weight of a given firm across different rankings. The results are derived as above, i.e. by simulating randomly the variation of all parameters jointly. High concentration of areas vertically denotes a stable ranking for that specific firm. Note that such behaviour tends to emerge both on the left and the right of the chart, whereas intermediate rankings are more volatile and erratic. Intuitively, this implies that both top and bottom companies are also more likely to respectively outperform and

Table 8: Average of the range and entropy assignments across various settings of $w \in [0, 1]$ for different specifications of the thresholds. The selected threshold settings used in the application are reported in bold.

Thresholds			Uncertainty Measures	
q	p	v	Entropy	Range
0	0.5	0.7	0.34	11.41
0	0.5	1	0.55	29.59
0.2	**0.6**	**0.8**	**0.34**	**11.08**
0.25	0.75	1.5	0.81	80.99
0.5	1	2	0.79	78.16

underperform competitors from a sustainability point of view companies under different scenarios.

Although the overall volatility of scores and rankings as a function of both thresholds and weights in some cases is not negligible, the results are nonetheless encouraging, since the choice of the parameters does not seem to overly affect the final ordering. As a consequence, we conclude that a neutral parameter setting could be a sweet spot in terms of model interpretability between the complexity of a hybrid MCDA approach and the subjectivity of expert opinions (Zavadskas et al., 2016) for deriving fine-tuned parameters.

Our choice of parameters is supported from a preliminary assessment from the present sensitivity analysis. Nonetheless, the selected settings are largely consistent with proposals in the literature. Among various strategies proposed so far, the rule-of-thumb approach of (Corazza et al., 2016) based on equal-weighting of criteria and either quantile-based or $|max - min|$ setting of thresholds are close to our proposal; other possibilities include expert elicitation (Doumpos and Figueira, 2019), hybrid methods aggregating opinions from expert opinion (Angilella and Mazzù, 2015) or, alternatively, endogenously determined parameters (Corazza et al., 2021).

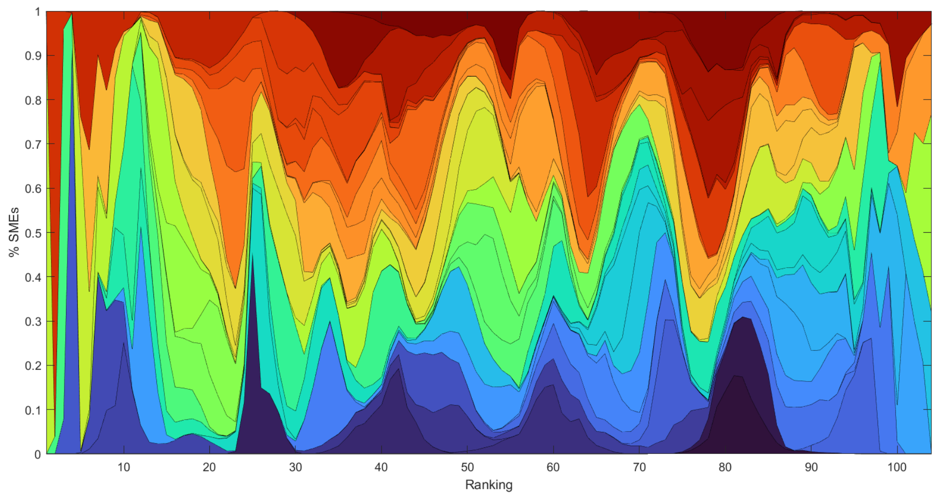

Figure 6: Distribution of firms' weights in the general standings across different settings. Note that on the x-axis the ranking for each company across $s = 10,000$ simulations is reported, while on the y-axis the percentage of all the SMEs for a given rank is represented with a color for each SME.

5.2.2 Assessing the robustness of the imputation procedure

In what follows, we briefly assess the impact of the proposed imputation procedure. Our point for making a further robustness check is that a replacement of missing values with a by-sector point estimate based on the worst value, might make pairwise comparisons between alternatives insensitive across different threshold settings. In this way, it could be hypothesized that the proposed approach might artificially induce robustness in the rankings and scores across different parameter settings.

Therefore, we generate replacements for missing values according to a multiple imputation procedure. We prudentially impute missing values by sampling respectively from a uniform, a normal, and a lognormal distribution 10,000 times. We assume prudentially that missing data are governed by distributions centered to the left of the observed data. Thus, we fit the above-mentioned distribution on observed values below the median.

A two-sample Kolmogorov-Smirnov test is applied to check whether each distribution of scores, whose missing values have been reconstructed with a multiple imputation method, is statistically different from the reference distribution of scores based on imputation with the worst sector value. The null hypothesis cannot be rejected for all the three tests. As for the normality of the distributions of scores, both the Jarque-Bera test and the Anderson-Darling test lead to a rejection of the null hypothesis at the 1% significance level.

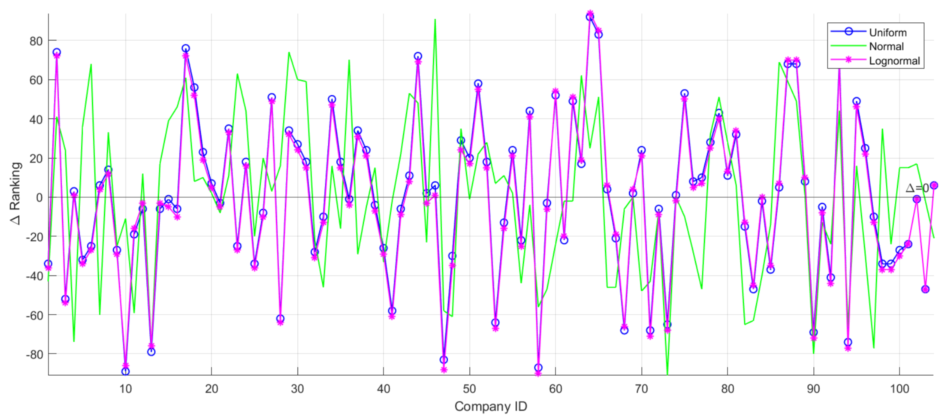

Figure 7: Each coloured plot denotes the differences between the reference ranking (see the horizontal line in black with $\Delta = 0$) and the rankings derived with multiple imputation methods.

Finally, in Figure 7 the difference between the reference ranking ($\Delta = 0$) and the rankings based on multiple imputation is reported, with notable differences for most companies, apart from a few exceptions for firms placed on the right tail of the distribution (see in particular IDs 8, 9, 11, 42, 56, 57, 70 and 91).

We conclude that, although the variability of the rankings is not negligible, the top firms are not only robust to different settings but also to various imputation methods.

5.3 Assessing the impact of a prudential single imputation procedure

As documented in Section 3, in this study, we make a relatively strong assumption with regard to the imputation of missing data. In accordance with the approximation that lack of data might imply unwillingness or inability to release information, especially due to direct and indirect costs borne by SMEs, we impute missing data in a given sector with the worst value of that sector, if available, otherwise are set equal to the global worst value. Such a heuristic replacement rule entails a variety of statistical and economic consequences. For a brief discussion of both we refer the reader to Section 3.1. However, here we further elaborate on this point, as we factor in other model-specific consequences. In particular, in Proposition 5.1 we formally show that, provided that some conditions are satisfied, and given a decision matrix $\mathbf{G}_{m\times n}$, whose elements g_{ij} contain the score for the alternative i and the criterion j, with exactly one missing value w.r.t. a criterion-alternative pair, it is always convenient for firm i to disclose the true (not publicly known) observation in place of the missing one, in the sense that the alternative is at least as good as if the value were unknown, according to the DM's preferences. Moreover, in Proposition 5.2 we extend proposition 5.1 to the case of two or more missing observations, for two or more alternatives, and we discuss why no conclusion can be actually reached in terms of the DM's preferences, since the "true" net flow depends also on pairwise comparisons between true unknown values.

Proposition 5.1 *Let a_i be the only alternative among the m ones for which its performance in relation to the j-th criterion is unknown (so, g_{ij} is prudentially set to 0), let $\tilde{g}_{ij} \geq 0$ be the unknown performance of a_i before transformation (1), and let $\tilde{\varphi}(a_i)$ be the net flow of a_i computed considering the true but unknown performance of a_i w.r.t. the j-th criterion. If there exist at least an alternative a_k, with $k \neq i$, such that*

- *$g_{kj} > 0$ in case the j-th criterion is to be maximized,*

- $g_{kj} < 1$ *in case the* j*-th criterion is to be minimized*

then $\varphi(a_i) \leq \tilde{\varphi}(a_i)$.

Proof. Let us focus on the case where the j-th criterion is to be maximized. Let us use the symbol tilde above the quantities of interest $(\tilde{\cdot})$ when they are related to the true but unknown values of those quantities.

The fact that $\tilde{g}_{ij} \geq 0$ implies the following regarding the local concordances and local discordances, for the pairs of alternatives (a_i, a_k) and (a_k, a_i) w.r.t. the j-th criterion:

- $C_j(a_i, a_k) \leq 1$ by definition, and $C_j(a_k, a_i) = 1$ since $g_{ij} - g_{kj} = -g_{kj} < 0 < q_j$ (see Equation (2));

- $\tilde{C}_j(a_i, a_k) \geq C_j(a_i, a_k)$ as $g_{kj} - \tilde{g}_{ij} \leq g_{kj} - g_{ij} = g_{kj}$ and $\tilde{C}_j(a_k, a_i) \leq C_j(a_k, a_i)$ since $\tilde{g}_{ij} - g_{kj} \geq g_{ij} - g_{kj} = -g_{kj}$ (see Equation (2));

- $D_j(a_i, a_k) \geq 0$ by definition, and $D_j(a_k, a_i) = 0$ since $g_{ij} - g_{kj} = -g_{kj} < 0 < p_j$ (see Equation (3));

- $\tilde{D}_j(a_i, a_k) \leq D_j(a_i, a_k)$ as $g_{kj} - \tilde{g}_{ij} \leq g_{kj} - g_{ij} = g_{kj} = g_{kj}$, and $\tilde{D}_j(a_k, a_i) \geq D_j(a_k, a_i)$ since $\tilde{g}_{ij} - g_{kj} \geq g_{ij} - g_{kj} = -g_{kj}$ (see Equation (3)).

So, concerning the global concordances of pairs (a_i, a_k) and (a_k, a_i), one has:

$$\tilde{C}(a_i, a_k) = \sum_{\substack{l=1 \\ l \neq j}}^{n} w_l C_l(a_i, a_k) + w_j \tilde{C}_j(a_i, a_k) \geq C(a_i, a_k) = \sum_{\substack{l=1 \\ l \neq j}}^{n} w_l C_l(a_i, a_k) + w_j C_j(a_i, a_k) \tag{9}$$

since $\tilde{C}_j(a_i,a_k) \geq C_j(a_i,a_k)$ and

$$\tilde{C}(a_k,a_i) = \sum_{\substack{l=1 \\ l \neq j}}^{n} w_l C_l(a_k,a_i) + w_j \tilde{C}_j(a_k,a_i) \leq C(a_k,a_i) = \sum_{\substack{l=1 \\ l \neq j}}^{n} w_l C_l(a_k,a_i) + w_j C_j(a_k,a_i) \tag{10}$$

as $\tilde{C}_j(a_k,a_i) \leq C_j(a_k,a_i)$.

Similarly, concerning the outranking indices of the pairs (a_i,a_k) and (a_k,a_i) one has:

$$\tilde{O}(a_i,a_k) \geq O(a_i,a_k) \tag{11}$$

since $\tilde{C}_j(a_i,a_k) \geq C_j(a_i,a_k)$ and $\tilde{D}_j(a_i,a_k) \leq D_j(a_i,a_k)$.

$$\tilde{O}(a_k,a_i) \leq O(a_k,a_i) \tag{12}$$

as $\tilde{C}_j(a_k,a_i) \leq C_j(a_k,a_i)$ and $\tilde{D}_j(a_k,a_i) \geq D_j(a_k,a_i)$.

Lastly, with reference to the net flows $\varphi(a_i) = \varphi^+(a_i) - \varphi^-(a_i)$ and $\tilde{\varphi}(a_i) = \tilde{\varphi}^+(a_i) - \tilde{\varphi}^-(a_i)$ (see Equation (6)), from the above one has that all the addends of $\varphi^+(a_i)$ are lower or equal than the corresponding addends of $\tilde{\varphi}^+(a_i)$ and that all the addends of $\varphi^-(a_i)$ are greater or equal than the corresponding addends of $\tilde{\varphi}^-(a_i)$, so:

$$\varphi^+(a_i) \leq \tilde{\varphi}^+(a_i) \quad \text{and} \quad \tilde{\varphi}^-(a_i) \geq \tilde{\varphi}^-(a_i) \tag{13}$$

therefore

$$\varphi(a_i) = \varphi^+(a_i) - \varphi^-(a_i) \leq \tilde{\varphi}(a_i) = \tilde{\varphi}^+(a_i) - \tilde{\varphi}^-(a_i) \tag{14}$$

This proves the thesis.

Similarly, one can prove the case where the j-th criterion is to be minimized. ■

Proposition 5.2 *Let $a_{i_1}, a_{i_2}, \ldots, a_{i_p}$, with $|\{i_1, i_2, \ldots, i_p\}| \in \{2, 3, \ldots, m-1\}$ be the alternatives among the m whose performance in relation to the j-th criterion are unknown (so, $g_{i_1 j}, g_{i_2 j}, \ldots, g_{i_p j}$ are prudentially set to 0), let $\tilde{g}_{i_1 j} \geq 0, \tilde{g}_{i_2 j} \geq 0, \ldots, \tilde{g}_{i_p j} \geq 0$ be the unknown performance of $a_{i_1}, a_{i_2}, \ldots, a_{i_p}$ respectively before transformation (1), and let $\tilde{\varphi}(a_{i_1}), \tilde{\varphi}(a_{i_1}), \ldots, \tilde{\varphi}(a_{i_p})$ be the net flows of $a_{i_1}, a_{i_2}, \ldots, a_{i_p}$ respectively computed considering the ture but unknown performance of $a_{i_1}, a_{i_2}, \ldots, a_{i_p}$ w.r.t. the j-th criterion. If there exist at least an alternative a_k with $k \notin \{i_1, \ldots, i_2, \ldots, i_p\}$ such that*

- *g_{kj} in case the j-th criterion is to be maximized,*
- *g_{kj} in case the j-th criterion is to be minimized,*

then $\varphi(a_i) \gtreqless \tilde{\varphi}(a_i)$.

Sketch of proof. Let us focus again on the case where the j-th criterion is to be maximized. Concerning the comparison of pairs of alternatives in which one alternative has an unknown performance and the other alternative has a known performance (as in Proposition 5.1), all the inequalities proved in Proposition 5.1 hold again. But as for the comparison of pairs of alternatives in which both the alternatives have unknown performance, none of the inequalities proved in Proposition 5.1 hold anymore. Therefore, the contributions of the net flows of the latter kind of comparisons cannot be further evaluated and, consequently, it is no longer possible to define an ordering between $\phi(a_i)$ and $\tilde{\phi}(a_i)$. Similarly, one can prove the case where the j-th criterion is to be minimized. ■

6 Conclusions

In this contribution we have analyzed the sustainability profiles in SMEs with a MCDA approach. We have set up a flexible model where a limited number of assumptions is necessary. Subsequently, we tackled the inherently unstructured nature of the problem by assuming relatively neutral preferences of the DM and that each pillar is equally significant for all sectors. Among others, our aim was to avoid relying either on hybrid approaches or subjective experts judgement. Furthermore, such approach ensured that ESG information is reported in a consistent way across all material topics.

Thus, we stressed the importance of identifying a relatively small set of variables that are relevant to firms and investors and to construct a set of (Global Reporting Initiative, 2022)-compliant criteria, in order to make the assessment of results aligned with widely accepted global standards for sustainability impact. Moreover, the constructed criteria displayed particularly low correlations. In this way, we ensured that in our scoring system all the SMEs converging to higher reporting standards were rewarded. The obtained rankings were aimed at capturing leader and laggard firms in terms of ESG performance, which was subsequently shown to be robust across different model parameterizations. Moreover, by adopting a prudential and intuitive imputation approach, we tailored our approach to the distinctive features of SMEs, in line with previous empirical contributions stressing that the size of a firm is positively related to the extent of voluntary disclosure (Prencipe, 2004). Since such imputation approach entails various economic and statistical consequence, we also proposed to assess how this might affect the firm's decision to actually release ESG data, under a conventional outranking MCDA approach, also in light of the above mentioned costs borne by SMEs.

A sensitivity analysis finally confirmed that our results are robust across different model parameterizations and imputation methods.

The quality of data represent a barrier that must be taken into account in research on small business finance. Nonetheless, further research should be extended towards three directions. First, more granular databases would allow to break down data by industry and by country, allowing to characterize in detail the firms' sustainability performance. Second, the integration of the ESG dimensions in a credit risk model for SMEs would unlock a deeper understanding of its influence on firms' credit worthiness. Finally, to shed light on key determinants of variations in ESG score, an extension of our model to panel data would allow to capture time-varying effects. An analysis of the joint influence of time, firm and country characteristics on ESG performance and its relationship with firms profitability is currently high on our agenda.

Disclaimer

This paper was developed within the project funded by Next Generation EU - "GRINS - Growing Resilient, INclusive and Sustainable" project (PE0000018), National Recovery and Resilience Plan (NRRP) – PE9 - Mission 4, C2, Intervention 1.3". The views and opinions expressed are only those of the authors and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Acknowledgments

We would like to thank participants at the 96th meeting of the EURO Working Group on Multiple Criteria Decision Aiding, ESSCA School of Management.

References

Altman, E. I. and Sabato, G. (2007). Modelling credit risk for SMEs: Evidence from the US market. *Abacus*, 43(3):332–357.

Altman, E. I., Sabato, G., and Wilson, N. (2010). The value of non-financial information in small and medium-sized enterprise risk management. *The Journal of Credit Risk*, 6(2):95.

Angilella, S. and Mazzù, S. (2015). The financing of innovative SMEs: A multicriteria credit rating model. *European Journal of Operational Research*, 244(2):540–554.

Baier, P., Berninger, M., and Kiesel, F. (2020). Environmental, social and governance reporting in annual reports: A textual analysis. *Financial Markets, Institutions & Instruments*, 29(3):93–118.

Bartolacci, F., Caputo, A., and Soverchia, M. (2020). Sustainability and financial performance of small and medium sized enterprises: A bibliometric and systematic literature review. *Business Strategy and the Environment*, 29(3):1297–1309.

Baumüller, J. and Grbenic, S. O. (2021). Moving from non-financial to sustainability reporting: Analyzing the EU commission's proposal for a Corporate Sustainability Reporting Directive (CSRD). *Facta Universitatis, Series: Economics and Organization*, (1):369–381.

Billio, M., Costola, M., Hristova, I., Latino, C., and Pelizzon, L. (2021). Inside the ESG ratings:(Dis) agreement and performance. *Corporate Social Responsibility and Environmental Management*, 28(5):1426–1445.

Brans, J.-P. and Vincke, P. (1985). Note—A Preference Ranking Organisation Method: (The PROMETHEE Method for Multiple Criteria Decision-Making). *Management Science*, 31(6):647–656.

Corazza, M., Fasano, G., Funari, S., and Gusso, R. (2021). MURAME parameter setting for creditworthiness evaluation: data-driven optimization. *Decisions in Economics and Finance*, 44(1):295–339.

Corazza, M., Funari, S., and Gusso, R. (2015). An evolutionary approach to preference disaggregation in a MURAME-based creditworthiness problem. *Applied Soft Computing*, 29:110–121.

Corazza, M., Funari, S., and Gusso, R. (2016). Creditworthiness evaluation of Italian SMEs at the beginning of the 2007–2008 crisis: An MCDA approach. *The North American Journal of Economics and Finance*, 38:1–26.

Czarnitzki, D. and Hottenrott, H. (2011). R&D investment and financing constraints of small and medium-sized firms. *Small Business Economics*, 36:65–83.

Deloitte (2023). EU Taxonomy and CSRD climate risks assessments.

Doumpos, M., Andriosopoulos, K., Galariotis, E., Makridou, G., and Zopounidis, C. (2017). Corporate failure prediction in the European energy sector: A multicriteria approach and the effect of country characteristics. *European Journal of Operational Research*, 262(1):347–360.

Doumpos, M. and Figueira, J. R. (2019). A multicriteria outranking approach for modeling corporate credit ratings: An application of the Electre Tri-nC method. *Omega*, 82:166–180.

D'Amato, A. (2020). Capital structure, debt maturity, and financial crisis: empirical evidence from SMEs. *Small Business Economics*, 55(4):919–941.

El Ghoul, S., Guedhami, O., Kwok, C. C., and Mishra, D. R. (2011). Does corporate social responsibility affect the cost of capital? *Journal of Banking & Finance*, 35(9):2388–2406.

Enders, C. K. (2022). *Applied missing data analysis.* Guilford Publications.

European Commission (2023). Corporate sustainability reporting.

Feng, Z. and Wu, Z. (2021). ESG disclosure, REIT debt financing and firm value. *The Journal of Real Estate Finance and Economics*, pages 1–35.

Figueira, J. R., Mousseau, V., and Roy, B. (2016). ELECTRE methods. In *Multiple criteria decision analysis: State of the art surveys*, International Series in Operations Research & Management Science, S. Greco, M. Ehrgott and J. Figuiera, pages 155–185. Springer.

García-Martínez, G., Guijarro, F., and Poyatos, J. A. (2019). Measuring the social responsibility of European companies: a goal programming approach. *International Transactions in Operational Research*, 26(3):1074–1095.

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

Gjergji, R., Vena, L., Sciascia, S., and Cortesi, A. (2021). The effects of environmental, social and governance disclosure on the cost of capital in small and medium enterprises: The role of family business status. *Business Strategy and the Environment*, 30(1):683–693.

Global Reporting Initiative (2022). Sustainability reporting guidelines.

Goletsis, Y., Psarras, J., and Samouilidis, J.-E. (2003). Project ranking in the Armenian energy sector using a multicriteria method for groups. *Annals of Operations Research*, 120:135–157.

Goyal, P., Rahman, Z., and Kazmi, A. A. (2013). Corporate sustainability performance and firm performance research: Literature review and future research agenda. *Management Decision*, 51(2):361–379.

Guney, Y., Hernandez-Perdomo, E., and Rocco, C. M. (2020). Does relative strength in corporate governance improve corporate performance? Empirical evidence using MCDA approach. *Journal of the Operational Research Society*, 71(10):1593–1618.

Höck, A., Klein, C., Landau, A., and Zwergel, B. (2020). The effect of environmental sustainability on credit risk. *Journal of Asset Management*, 21(2):85–93.

Hwang, C.-L., Yoon, K., Hwang, C.-L., and Yoon, K. (1981). Methods for multiple attribute decision making. *Multiple attribute decision making: methods and applications a state-of-the-art survey*, pages 58–191.

Keeney, R. L. and Raiffa, H. (1993). *Decisions with multiple objectives: preferences and value trade-offs.* Cambridge University Press.

Khan, M., Serafeim, G., and Yoon, A. (2016). Corporate sustainability: First evidence on materiality. *The Accounting Review*, 91(6):1697–1724.

Kotsantonis, S. and Serafeim, G. (2019). Four things no one will tell you about ESG data. *Journal of Applied Corporate Finance*, 31(2):50–58.

Lindsey, L. A., Pruitt, S., and Schiller, C. (2022). The cost of ESG Investing. *Available at SSRN 3975077.*

Malesios, C., De, D., Moursellas, A., Dey, P. K., and Evangelinos, K. (2021). Sustainability performance analysis of small and medium sized enterprises: Criteria, methods and framework. *Socio-Economic Planning Sciences*, 75:100993.

Moro, A. and Fink, M. (2013). Loan managers' trust and credit access for SMEs. *Journal of Banking & Finance*, 37(3):927–936.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., and Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, 79:214–232.

Prencipe, A. (2004). Proprietary costs and determinants of voluntary segment disclosure: evidence from Italian listed companies. *European Accounting Review*, 13(2):319–340.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, 53:49–57.

Rodríguez-Gutiérrez, P., Guerrero-Baena, M. D., Luque-Vílchez, M., and Castilla-Polo, F. (2021). An approach to using the best-worst method for supporting sustainability reporting decision-making in SMEs. *Journal of Environmental Planning and Management*, 64(14):2618–2640.

Roy, P. K. and Shaw, K. (2021). Developing a multi-criteria sustainable credit score system using fuzzy BWM and fuzzy TOPSIS. *Environment, Development and Sustainability*, pages 1–32.

Saaty, T. L. (1988). *What is the analytic hierarchy process?* Springer.

Sahin, Ö., Bax, K., Czado, C., and Paterlini, S. (2022). Environmental, Social, Governance scores and the Missing pillar—Why does missing information matter? *Corporate Social Responsibility and Environmental Management*, 29(5):1782–1798.

Shehata, N., Salhin, A., and El-Helaly, M. (2017). Board diversity and firm performance: evidence from the UK SMEs. *Applied Economics*, 49(48):4817–4832.

Siddiqi, N. (2017). *Intelligent credit scoring: Building and implementing better credit risk scorecards*. John Wiley & Sons.

Sorensen, E., Chen, M., and Mussalli, G. (2021). The quantitative approach for sustainable investing. *The Journal of Portfolio Management*, 47(8):38–49.

Torelli, R., Balluchi, F., and Furlotti, K. (2020). The materiality assessment and stakeholder engagement: A content analysis of sustainability reports. *Corporate Social Responsibility and Environmental Management*, 27(2):470–484.

Tsikriktsis, N. (2005). A review of techniques for treating missing data in OM survey research. *Journal of Operations Management*, 24(1):53–62.

Voulgaris, F., Doumpos, M., and Zopounidis, C. (2000). On the evaluation of Greek industrial SME's performance via multicriteria analysis of financial ratios. *Small Business Economics*, 15:127–136.

Zavadskas, E. K., Govindan, K., Antucheviciene, J., and Turskis, Z. (2016). Hybrid multiple criteria decision-making methods: A review of applications for sustainability issues. *Economic research-Ekonomska istraživanja*, 29(1):857–887.

Zionts, S. (1979). MCDM—if not a roman numeral, then what? *Interfaces*, 9(4):94–101.