PHILIPPE DUPUY
Grenoble Ecole de Management – 38000 Grenoble, France.

JEAN-CHARLES GARIBAL
Grenoble Ecole de Management – 38000 Grenoble, France.

MATHIEU JOUBREL
ValueCo – Paris, France.

Mitigating company characteristics as external biases in ESG valuation models

Abstract

We study the impact of firms' characteristics such as firms' size, sector, and geographical activity area in ESG scorings. These characteristics may create biases in the scoring models. For example, larger firms may show on average better ESG scores than smaller firms. We propose a methodology to mitigate these biases and compute a score that is free from any factors including firms' characteristics. We also propose a framework to detect outperforming companies regardless of any factors. Our methodology can break down final ESG scores into a component related to these external factors and a component unrelated to them.

Keywords: *ESG, sustainability, factors, scoring.*

JEL classification: *C45, C53, C58, G01, G11.*

I. Introduction

Socially Responsible Investments (SRI) has been one of the key drivers of the finance industry over the past few years. From a few pioneers 15 years ago, 4000 financial institutions are now signatories of the United Nations' Principles for Responsible investment (UN PRI) and the total assets managed by these signatories reached in 2022 more than 120T$. This trend proves the growing commitment of the finance industry to ESG and responsible investment. However, the rapidly evolving ESG landscape and growing market demand for ESG products entail new risks for all stakeholders of the ecosystem. Especially, standards, frameworks, and initiatives have proliferated to the extent that it has become exceedingly difficult to effectively monitor and compare investors' claims about their compliance with ESG investing. For instance, many asset managers have developed internal proprietary scoring models to justify their ESG commitment. But these models can hardly be compared from one to the other.

As ESG scores available from rating agencies already show a low correlation (Chatterji et al., 2015), the multiplication of proprietary models is likely to amplify this phenomenon. The raw data underlying these scores also still lacks the quality to support efficient decision-making. Classic issues or divergences in ESG ratings for designers of ESG scoring models include data availability, transparency and update frequency, as well as the lack of standardization (Dupuy and Garibal, 2022). Furthermore, additional internal biases come from divergences in the raters' scoring methodologies. Berg et al. (2020) find three sources of divergences among raters: different scope of categories, different measurement of categories, and different weights of categories. The first divergence is in the scope, i.e., raters do not define the phenomenon to measure in the same way. The second divergence is in measurement, i.e., the lack of a consensual proxy to measure a chosen indicator (biodiversity, happiness...). Finally, the third divergence is in the way raters aggregate the indicators; since each rater decides independently which aspects of the ESG scoring deserve more weight.

In addition to the biases listed above, ESG scores can suffer from external biases, or biases that are not directly linked to issuers' sustainability performance, disclosure, or risk exposure. These biases can stem among others from factors related to firms' characteristics such as size, sector and geographical activity area. Yet, it is well known that taking these biases into account can improve investment decisions (Amel-Zadeh and Serafeim, 2018, Ratsimiveh et al., 2020 and Berk and van Binsbergen, 2021). For instance, firms' size and firms' sustainability performance seem to be related (Drempetic et al.,2019; Taliento et al.,2019) since larger companies tend to profit from economies of scale when addressing ESG challenges (OECD, 2022; Bissoondoyal-Bheenick et al., 2023). The industry sector may have an impact on ESG performance as well since each sector is not exposed to the same ESG risks (Taliento et al., 2019). Especially, it is well known that higher ESG scores are associated with higher diversification in industry sectors (Barros et al., 2024).

It is important to note that ESG rankings should ideally reflect firms' intrinsic sustainability performance, rather than being driven by external characteristics such as size, sector, or country. Regulatory environments, cultural norms, and market constraints often vary across countries and industries, but these differences do not necessarily reflect a firm's strategic commitment to sustainability. Moreover, the availability of ESG data is frequently influenced by a firm's size and geographic location (OECD, 2025). These external factors can introduce systematic biases into ESG classifications, which may, in turn, distort asset pricing and the allocation of capital.

In this context, the aim of this paper is to detect and quantify ESG scoring models biases arising from the above-mentioned external factors (i.e., firms' size, activity sector and geographical area). We also provide tools to measure the absolute biases of these models. We test our methodology on a set of ESG models provided by asset managers and gathered by a fintech called ValueCo. Our methodology holds the promise of refining asset allocation decisions by removing external biases from ESG scores, potentially yielding signals that align more closely with intrinsic performance. Finally, we show how external biases can even be used to enhance firm-level analysis instead of merely biasing investment decisions.

Crucially, systematic external biases are not merely measurement issues; if ESG scores consistently favor large firms or certain regions, they can lead to persistent asset mispricing and skew capital allocation, ultimately hindering the efficient identification of positive-NPV projects (Goss & Roberts, 2011; Krüger, 2015).

The remaining of this paper is organized as follows. Section 2 describes the data while Section 3 highlights the biases. Section 4 is dedicated to the mitigation of biases. Section 5 concludes.

II. Data description

The results presented in this paper are based on the cross-sectional analysis of a set of 9089 listed companies of various countries and sectors in 2024. These companies are well diversified in terms of size, sector and geographical origins and exposures (See Annex A). We aggregate ESG scores developed internally by asset managers to build a realistic market view of these listed equities. This dataset is extracted as of January 2024 from the proprietary database of ValueCo, a fintech collecting proprietary sustainability-related scores developed in-house by professional investors: asset managers, private banks, institutional investors. ValueCo has collected the ratings from 64 proprietary ESG rating models developed by 45 asset managers in Europe and North America managing a total of \$8.9 trillion in 2024. Although this figure represents only a subset of the approximately \$128 trillion total market capitalization of listed equities (Statista, 2025), the diversity of approaches and strategies employed by the contributing asset managers helps mitigate concerns regarding the representativeness of our sample. (See Annex B).

As confirmation, the average pairwise correlation among our 64 proprietary ESG scoring models is only 38%, lower than the 54% reported by Berg et al. (2020) for traditional rating agencies. This lower correlation highlights the fragmented nature of current ESG assessments and reinforces the need for bias-correction methodologies.

We preprocess the scores to mitigate issues arising from differences in scale or distribution across the 64 ESG scoring models. We normalize the output of each model on the investment universe using z-scores, ensuring that each model's average score is 0 and its standard deviation is 1. This makes the dispersion of scores from different models directly comparable and reduces the risk of misleading under- or over-scorings. We call the median score, obtained from the 64 scoring models for each firm, the ESG score consensus of the company.

III. Empirical Analysis: highlighting the bias

In Figure 1, we start by showing that the ESG score consensus exhibits external biases by box plotting it on different subsamples of the investment universe. In Figure 1, the horizontal line represents the mean of the distribution, the two boxes are the first and third quartiles and the vertical whiskers indicate the total range of the ESG score consensus in each activity sector (Panel A), geographic area (Panel B) and company size (Panel C). Any values outside the whiskers (computed as a function of the interquartile range) are represented with an individual point and considered as outliers.

In Panel A of Figure 1, we observe that the average values are different from one sector to the other. Some sectors have the same average as the investment universe (Consumer discretionary, Materials) while others show over half of a standard deviation of difference with it (Energy, Health care). The interquartile ranges vary between about 0.5 and 2 standard deviations, while the total range remains consistent across the different sectors. Considering that about two thirds of the model owners integrated into the ESG score consensus adopt a best-in-class strategy, their assessment of activity sectors must be balanced and as comparable as possible.

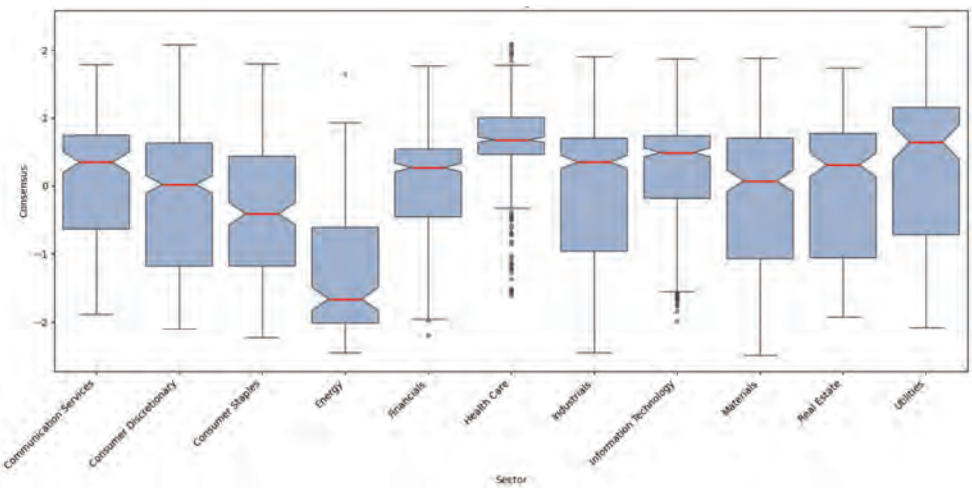
We now visualize the same plot using the most represented countries and geographical areas (Panel B). Similarly, we can observe significant differences in the distribution of the most represented geographical areas. These differences are also critical to consider. Indeed, they may result from ESG dumping from companies in wealthier countries compared with companies in poorer and less regulated ones. These differences may also come from the share of each activity sector in each country. When a model is biased towards activity sectors, it may add to the bias towards geographical areas too. Hence, reducing biases towards activity sectors mechanically reduces biases towards geographical areas.

Finally, we plot the normalized ESG scores consensus distribution by company size (Panel C). We use the market value as of 01/01/2025 as a proxy to create groups of companies of varied sizes. As ESG models mostly rely on disclosed corporate data, the company size factor is often a source of biases in ESG scores. Larger companies tend to have more resources to dedicate to marketing, reporting, and investors' relations. Although some of them exhibit good ESG performances, the overall better scores of larger companies can partially be linked to their greater disclosure capabilities. An efficient model should be able to sort the communication efforts from the actual effects of the company's CSR policies. The ESG score consensus seems balanced on small, mid, and large cap companies, but micro and nano cap companies stand out with lower average ratings, which is consistent with theory.

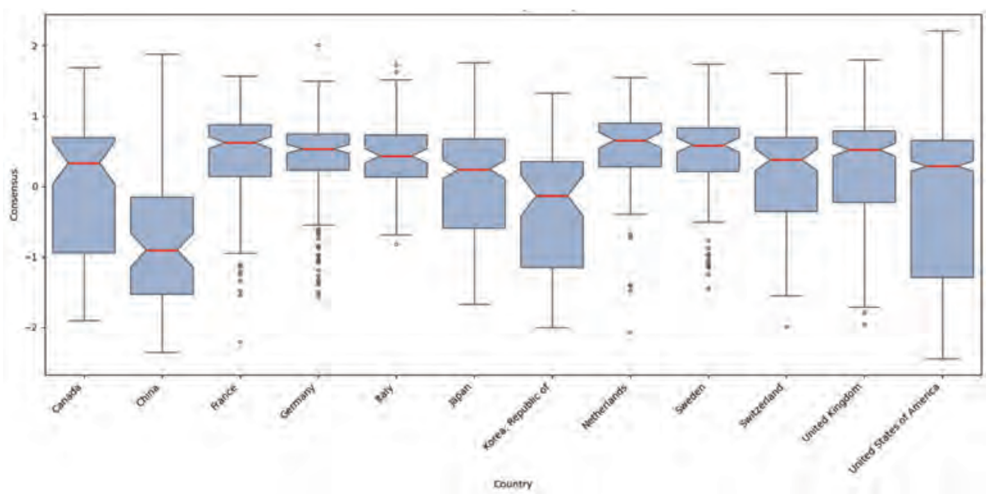
Figure 1. ESG score consensus by sector, geographical area, and company size"

Note: Consensus ESG scores (median scores) by sector, location and company size. The horizontal line represents the mean of the distribution, the two boxes are the first and third quartiles and the vertical whiskers indicate the total range of the ESG score consensus in each activity sector (Panel A), geographic area (Panel B) and company size (Panel C).

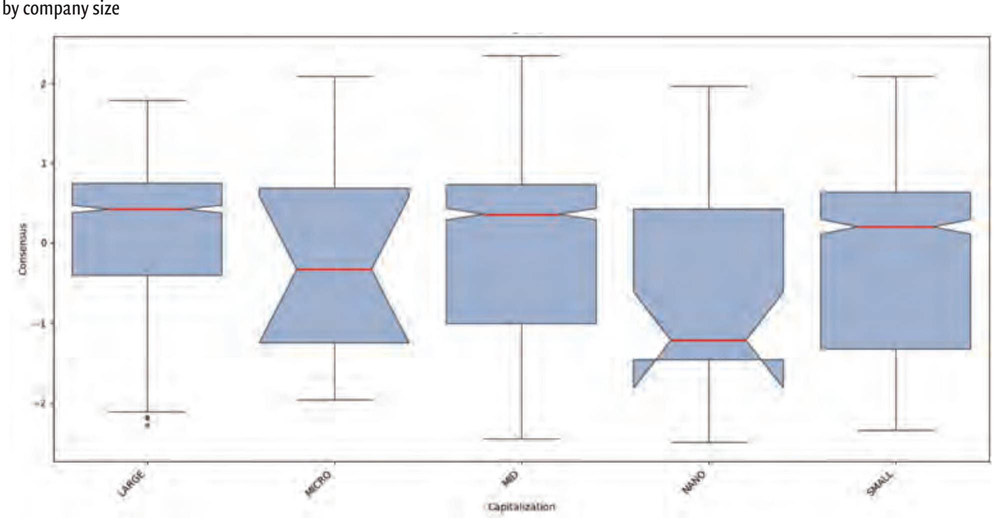
Panel A: by sector



Panel B: by geographical area



Panel C: by company size



We showed that the ESG score consensus is biased towards three external variables: activity sector, geographical area and firms' size. Although there may be many other factors towards which the consensus is biased, these three factors are widely recognized as bias vectors which can significantly affect investment decisions (Ratsimiveh et al., 2020; Berk and van Binsbergen, 2021). Yet an efficient model should be well-balanced regarding these three factors specifically. Consequently, we suggest in the next sections solutions to mitigate and control these biases in a bid to improve the ESG score consensus relevance.

IV. Biases offsetting

We attempt to mitigate the biases linked to external factors by performing a regression using only these three factors as input variables. The aim is to isolate the contribution of these factors to remove them from the ESG score consensus, as suggested by Ratsimiveh et al. (2020).

However, a major hypothesis to extract the individual contribution of each external factor is their mutual independence. As already mentioned in this paper, biases are interconnected, which is likely to modify the results and leads to inefficient Ordinary Least Squares (OLS) estimators (Wold et al., 2001). Therefore, we use as regressor a Partial Least Square (PLS) regressor. This regressor first aggregates the input variables into independent variables before fitting multiple linear regressors on this new dataset. The PLS regressor only relies on three variables, the three external factors, to explain the variations of the ESG score consensus. Supposing the number of observations and variables are large, PLS is asymptotically normal for the "best" forecast implied by a linear latent factor model. In stock market data, PLS shows relatively accurate out-of-sample forecasts of returns and cash-flow growth (Kelly and Pruitt, 2013; 2015).

First, recall the Multivariate Linear Model (MLR) with $\boldsymbol{X}$ a $(n \times m)$ matrix of explicative variables (the three external factors) and $\boldsymbol{S}$ a $(n \times p)$ matrix of targeted variables (the ESG scores consensus):

$$\boldsymbol{S} = \boldsymbol{X\beta} + \varepsilon_X \quad (1)$$

with β the traditional OLS vector of slope parameters considered as being the best linear unbiased estimator. The idea of using PLS instead of OLS is simply related to the inefficiency of the OLS estimator when there is multicollinearity among the variables in $\boldsymbol{X}$. The underlying model of multivariate PLS regression, with $l \leq m$ components, is:

$$\begin{cases} \boldsymbol{X} = \boldsymbol{TP}' + \varepsilon_T \\ \boldsymbol{S} = \boldsymbol{UQ}' + \varepsilon_U \end{cases} \quad (2)$$

with $\boldsymbol{T}$ and $\boldsymbol{U}$ two $(n \times l)$ matrices that are the projections of $\boldsymbol{X}$ and $\boldsymbol{S}$ respectively, and $\boldsymbol{P}$ and $\boldsymbol{Q}$ their respective loading matrices of dimensions $(m \times l)$ and $(p \times l)$ respectively. Finally, ε_T and ε_U are their respective error terms. $\boldsymbol{T}$ and $\boldsymbol{U}$ are obtained by maximizing their covariance when maximizing the covariance between $\boldsymbol{X}$ and $\boldsymbol{S}$ decompositions:

$$\max_{\boldsymbol{p}_j, \boldsymbol{q}_j} \mathrm{E}[(\boldsymbol{p}_j \boldsymbol{X}) \times (\boldsymbol{q}_j \boldsymbol{S})] \quad (3)$$

with $\boldsymbol{p}_j$, $\boldsymbol{q}_j$ the j-th vectors, with $j = 1, \ldots, l$, contained in $\boldsymbol{P}$ and $\boldsymbol{Q}$ respectively, so that $t_j = p_j X$ and $u_j = q_j S$, the j-th vectors of $\boldsymbol{T}$ and $\boldsymbol{U}$ respectively. Then, $\hat{\boldsymbol{S}} = \boldsymbol{UQ}'$. Both the original ESG scores consensus $\boldsymbol{S}$ and the external ESG score consensus $\hat{\boldsymbol{S}}$ are normalized for the analyses $(m \times l)$.

Maximizing the covariance between X and S decompositions through latent factors, T and U, is close to Principal Component Analysis (PCA) in maximizing the variance of the data into a set of orthogonal factors, i.e., the Principal Components (PCs). However, PCA is not supervised as PLS is, i.e., PCA does not simultaneously maximize the covariance between T and U, and then, between X and S decompositions (Zou et al., 2006).

Performing this regression, we obtain an adjusted R^2 of 31.6% and a Mean Squared Errors (MSE) of 0.684. Figure 2 shows the predicted scores $\hat{S}$ against the actual ESG score consensus S. We observe that most companies are either in the top right or in the bottom left corner, showing that they have an ESG score consensus aligned with their external factors. Most of the companies in the top left or bottom right corners remain close to the x = y identity line, meaning only a few of them are perceived differently from what their external factors would suggest.

To extract the biases linked to these factors from the ESG scores consensus, we study the differences between the predictions of the PLS regressor $\hat{S}$ and the targeted ESG score consensus S, also called residuals. To simplify the notations, let us define $\varepsilon_U = \varepsilon$ so that:

$$\boldsymbol{\varepsilon_U} = \boldsymbol{\varepsilon} = \hat{\mathbf{S}} - \mathbf{S} \quad (4)$$

The residuals represent the part of the score that cannot be explained by the input values. All the information contained in the residuals can only be explained by factors that are independent of these input values. It means all the biases linked to the external factors should be removed from the residual scores ε, only keeping the unbiased part of the ESG score consensus. We, therefore, normalize the vector of residual scores, ε, we just obtained to compute a normalized residual score $\sim S_i$ for firm i as follows:

$$\tilde{S}_i = \frac{\varepsilon_i - \mu_{\varepsilon_i}}{\sigma_{\varepsilon_i}} \quad (5)$$

This new score $\tilde{\boldsymbol{S}}$ represents the ESG performances of the companies independently of the three external factors and is called the residual ESG scores consensus. As we did for both the original ESG scores consensus $\boldsymbol{S}$ and the external ESG score consensus $\hat{\boldsymbol{S}}$, $\boldsymbol{S}$ is also normalized for the analyses.

We computed two new scores: i) a score recomposed from the regression of the ESG score on the three external factors $\hat{\boldsymbol{S}}$; and ii) a residual score that is independent from these factors . The first score represents the part of the ESG score consensus explained by the external factors and the second explains the rest of the score independently from these same external factors.

Figure 3 shows the improvement of the residual ESG score consensus compared to the normalized ESG scores consensus. Even though the different sectors are still not perfectly equivalent, we notice an improvement in the dispersion of ESG scores. We observe similar results in the country and company size segmentation. Although such biases are attenuated, they persist, suggesting that the residual ESG score consensus represents a refined version of the original consensus-one that is more independent from exogenous external factors. The remaining biases could be further mitigated with a better fit between the external factors and the ESG scores consensus.

Figure 2. External ESG scores consensus versus ESG scores consensus

Note: Consensus ESG scores (median score) and external ESG scores consensus. ESG scores consensus are demeaned and reduced to exhibit a zero mean and unit variance.

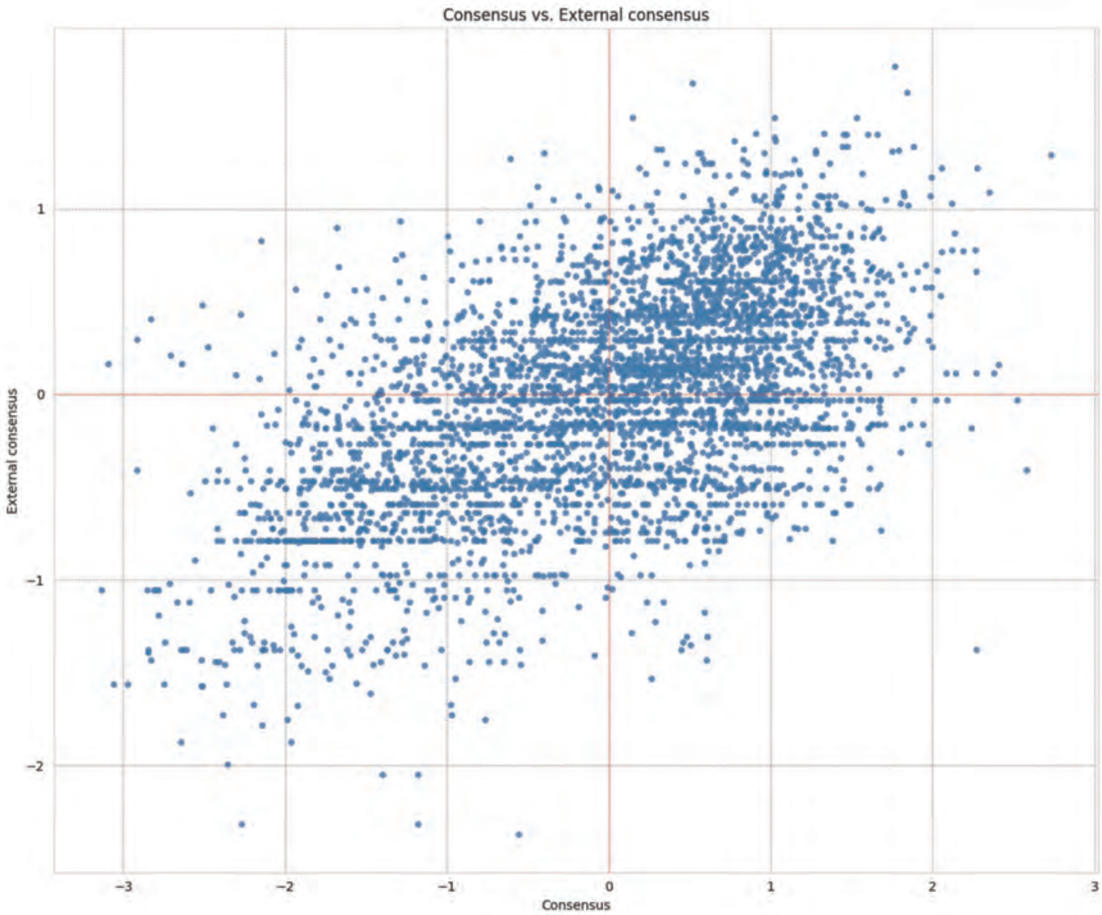
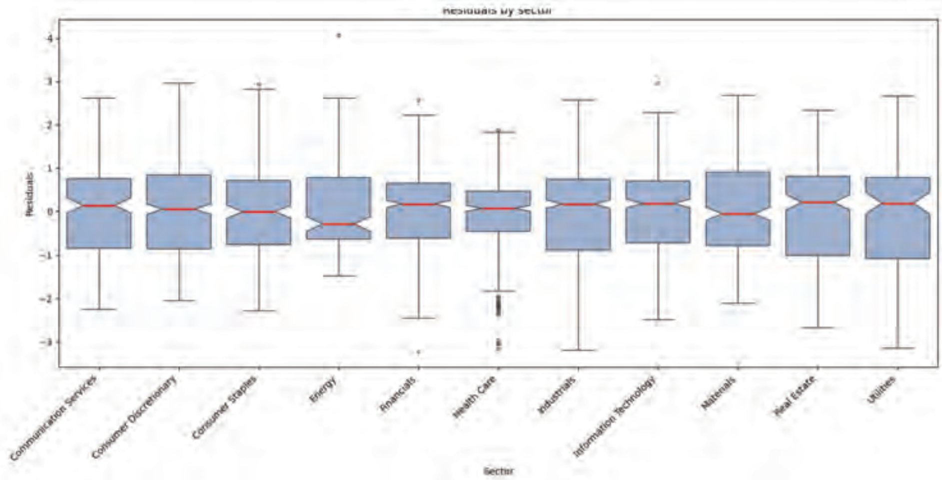


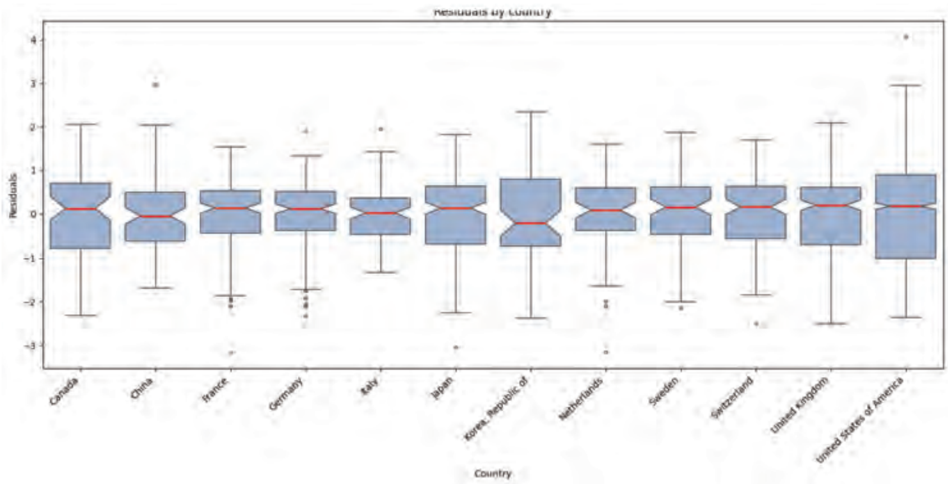
Figure 3. Residual ESG scores consensus

Note: Residual ESG scores consensus by sector, location and company size. The horizontal line represents the mean of the distribution, the two boxes are the first and third quartiles and the vertical whiskers indicate the total range of the ESG score consensus in each activity sector (Panel A), geographic area (Panel B) and company size (Panel C).

Panel A: by sector



Panel B: by geographical area



Panel C: by company size

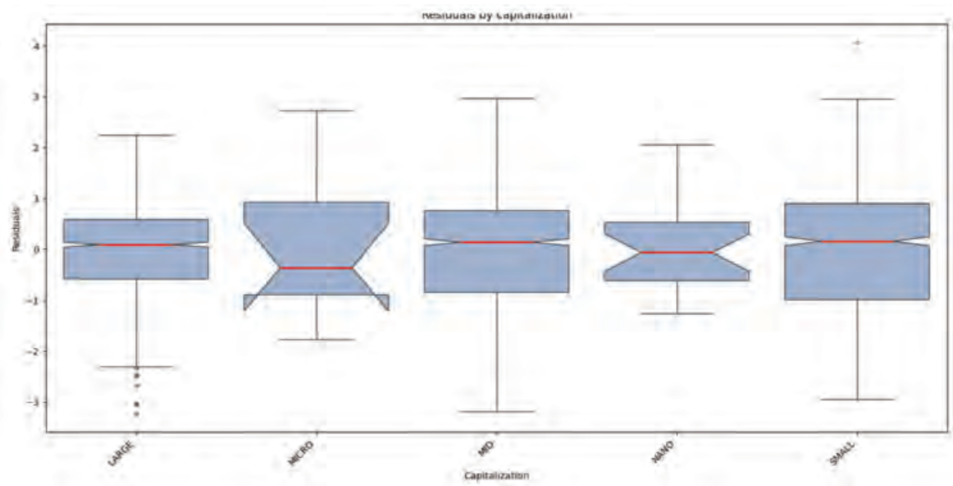


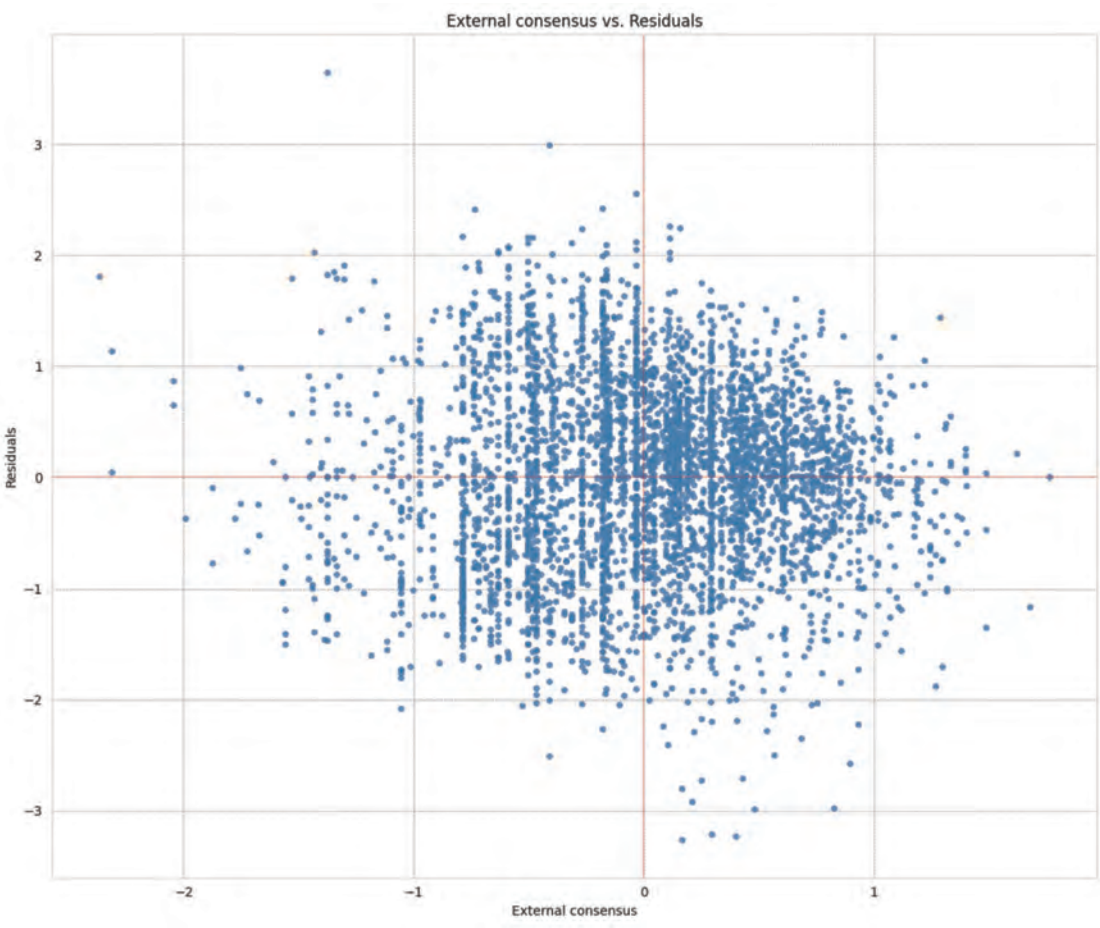
Figure 4 shows the relationship between the two new scores $\hat{S}$ and $\tilde{S}$, by plotting the residual scores (Y-axis) against the external factors-based scores (X-axis). Each point represents a company from the investment universe. The residual score is on the Y-axis, so the higher the company is on the plot the higher its ESG score (as explained independently from the external factors). The external factors-based score is on the X-axis, so the more on the right the company is, the more advantaged by its external factors it is.

In Figure 4, we can distinguish four main groups of companies:

The first group is in the top right corner: these companies belong to activity sectors, geographical areas and they exhibit sizes that favor higher ESG scores. They do have good intrinsic ESG performances as evaluated by the ESG score consensus. As an example, ERG S.p.A. (IT0001157020) is a major Italian energy firm in Europe's utilities sector. Its external consensus of +1.03 indicates a favorable operating environment which typically drives ESG scores higher. With a residual score of +1.09, the company even slightly outperforms these expectations, affirming its strong intrinsic sustainability practices relative to its peers.

Figure 4. Residual ESG scores consensus versus external ESG scores consensus

Note: Residual and external ESG scores consensus. ESG scores are demeaned and reduced to have a zero mean and unit variance.



The second group is in the bottom left corner: these companies are in activity sectors, geographical areas and they exhibit sizes that favor low ESG scores. They show poor intrinsic ESG performances as evaluated by the ESG score consensus. One of these firms is Kweichow Moutai Co. (CNE0000018R8), a Chinese beverage company. Its external consensus is –0.92, and its residual score of –1.23 still indicates overall below-expected intrinsic performance in 2024 according to the consensus of asset managers. This alignment of both scores suggests that, in an environment that already discounts ESG potential, the company would struggle to elevate its sustainable practices above industry norms.

The third group is in the top left corner: these companies belong to activity sectors, geographical areas and they exhibit sizes that favor low ESG scores, but they still show good intrinsic ESG performances as evaluated by the ESG score consensus. An example of such firm is China Vanke Co. (CNE0000000T2), a prominent Chinese real estate developer. Despite facing an external consensus of –0.82 reflecting less supportive market conditions, its robust residual score of +1.17 signals strong intrinsic ESG performance. This divergence indicates that, although its external environment limits typical ESG ratings, China Vanke's internal sustainability efforts allow it to substantially outperform expectations, setting it apart from similar peers.

The fourth group is in the bottom right corner: these companies belong to activity sectors, geographical areas and they exhibit sizes that favor high ESG scores, but still they have poor intrinsic ESG performances as evaluated by the ESG score consensus. For instance, BICO Group AB (SE0013647385), a biotechnology company headquartered in Sweden, operates in a sector and region that generally foster high ESG ratings, as reflected by an external consensus of +0.96. However, its residual score of –1.24 reveals that the company's intrinsic ESG performance lags behind these external expectations. This underperformance suggests that despite favorable external factors, BICO's internal sustainability practices may need improvement.

The proposed split of the ESG scores consensus into residual and external scores, allows for a more advanced analysis of each company's performance. Pedersen and Fitzgibbons (2020) suggest that this dual view would help ESG scores play their main roles: assessing a company's intrinsic performance and driving the investor's preferences.

While designing a bias-free ESG scoring model is impossible (due, for instance, to differences in disclosure requirements between countries), this methodology shows what part of the ESG score can be imputed to the three external factors. Once the individual performances of the companies are isolated, investors following best-in-class strategies can improve decisions on their investment universe while keeping in mind the highest-scoring sectors, sizes and geographical areas.

The proposed methodology is efficient in the breakdown of the impact of each factor when the ESG score consensus, $\boldsymbol{S}$, $\hat{\boldsymbol{S}}$ and $\tilde{\boldsymbol{S}}$ are statistically different from each other. Table 1 shows the results when testing whether the ESG scores, $\boldsymbol{S}$, $\hat{\boldsymbol{S}}$ and $\tilde{\boldsymbol{S}}$, are different from each other, based on the Wilcoxon-Mann-Whitney test (Wilcoxon, 1945; Mann and Whitney, 1947). Results suggest that the ESG scores $\boldsymbol{S}$, $\hat{\boldsymbol{S}}$ and $\tilde{\boldsymbol{S}}$ are different from each other according to the values of their test-statistics and their associated p values for a risk level $\alpha = 5\%$. Indeed, testing if S and $\hat{S}$ are different leads to a p value of $0.00 < \alpha = 5\%$, resulting to the non-rejection of the alternative hypothesis: "the two tested scores are statistically different". The same appears when comparing S with $\tilde{S}$ and $\hat{S}$ with $\tilde{S}$, resulting in p values equal to 0.19% and 3.6% respectively.

Table 1. Difference in ESG scores

Note: differences across the three ESG score consensus $\mathbf{S}$, $\hat{S}$ and $\tilde{S}$, are tested using Wilcoxon-Mann-Whitney test. In the first row, w, expressed in 10^5, refers to the value of the test-statistic associated to the p value in the next row below. p values are compared to risk level $\alpha = 5\%$ in the next row. True refers to the non-rejection of the alternative hypothesis: "the two tested scores are statistically different", while False refers to the non-rejection of the initial hypothesis: "the two tested scores are not statistically different". The last row gives the correlation ρ between each pair of scores.

	S **vs** $\hat{S}$	S **vs** $\tilde{S}$	$\hat{S}$ **vs** $\tilde{S}$
w	89.25	82.67	77.32
p value (in %)	0.00	0.19	3.6
Is significant:	True	True	True
ρ	53%	85%	0%

Correlations denoted ρ between $\boldsymbol{S}$, $\hat{\boldsymbol{S}}$, and $\tilde{\boldsymbol{S}}$ are displayed in the last row of Table 1. The high correlation (85%) between $\hat{\boldsymbol{S}}$ and $\boldsymbol{S}$ highlights the strong alignment between intrinsic performance and overall ESG evaluations, suggesting that the consensus retains substantial informational value beyond external factors. However, the lack of correlation (0%) between $\hat{\boldsymbol{S}}$ and $\tilde{\boldsymbol{S}}$ validates the effectiveness of the residual methodology in eliminating biases from external variables. The moderate correlation (53%) between $\boldsymbol{S}$ and $\hat{\boldsymbol{S}}$ signals the significant—but not overwhelming-role of external factors in shaping ESG scores.

For asset managers, these findings stress the importance of differentiating intrinsic performance from biases to enhance portfolio alignment with sustainable investment goals. By using residual-based scores, they can identify high-performing firms regardless of their size or sector to fair capital allocation and foster sustainability.

V. Conclusion

The wide variety of ESG scores on the market is the consequence of both the low quality of ESG data and the variety of views among investors. It is legitimate to have different understanding on ESG, but not to disagree on the very definition of indicators and measurement methods (Amel-Zadeh and Serafeim, 2018) as well as on the standardization of the indicators (Dupuy and Garibal, 2022). In this paper, we aim to detect and quantify the biases of the ESG score consensus from external factors. We choose to focus on three external factors: activity sector, company size and geographical area. These factors are widely recognized as bias vectors and can significantly influence investment decisions (Ratsimiveh et al., 2020; Berk and van Binsbergen, 2021).

We rely on PLS regression to create two independent scores. They represent the part of the ESG score consensus explained by the external factors alone and the part explained by the actual intrinsic performances of the company. We achieve a lower dependence of their ESG scores relative to external factors with a correlation of 53% between external and original ESG consensus, against 85% of correlation between residual and original ESG consensus. Furthermore, we propose a dual view of the original ESG score consensus showing how the companies of the investment universe can outperform or not their peers, as defined by the external factors.

Our results can be useful to ESG analysts, as bias analysis is mandatory to check whether the performances of an ESG scoring model matches their investment strategy. We provide tools to evaluate the absolute biases of any models. For investors, we show how the external biases arising from their ESG models can be used to enhance their firm-level analysis instead of merely biasing their investment decisions.

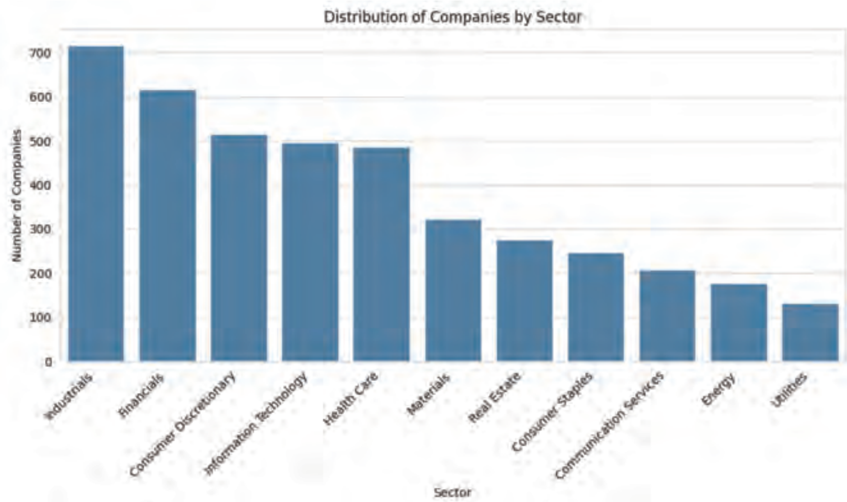
Further research could be done into assessing the impact of different factors on different markets to identify the factor structure across different markets. This will help to go deeper in the analysis of the divergences among proprietary ESG models.

References

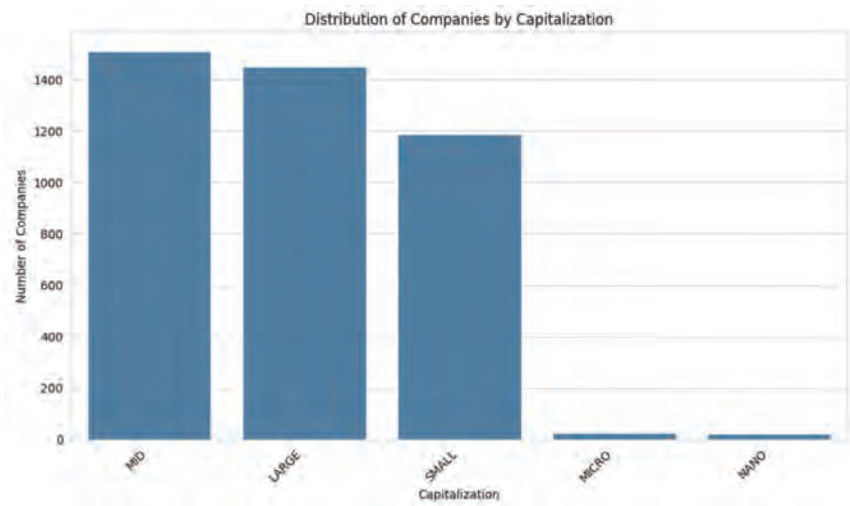
- Amel-Zadeh, A., Serafeim, G., 2018. Why and How Investors Use ESG Information: Evidence from a Global Survey. Financial Analysts Journal, 74 (3), 87–103.
- Barros, V., Matos, P., Sarmento, J., Vieira, P., 2024. ESG performance and firms' business and geographical diversification: An empirical approach. Journal of Business Research 172, 114392.
- Berg, F., and Kölbel, J., Rigobon, R., 2020. Aggregate Confusion: The Divergence of ESG Ratings. http://dx.doi.org/10.2139/ssrn.3438533
- Berg, F., Kölbel, J., Pavlova, A., Rigobon, R., 2021. ESG Confusion and Stock Returns: Tackling the Problem of Noise. http://dx.doi.org/10.2139/ssrn.3941514
- Berk, J., B., and van Binsbergen, J., H., The Impact of Impact Investing (August 21, 2021), https://ssrn.com/abstract=3909166
- Bissoondoyal-Bheenick, E., Brooks, R., Do, H., X., 2023. ESG and firm performance: The role of size and media channels. Economic Modelling 121, 106203.
- Chatterji, A., Durand, R., Levine, D., Touboul, S., 2015. Do ratings of firms converge? Implications for managers, investors and strategy researchers. Forthcoming, Strategic Management Journal. HEC Paris Research Paper No. SPE-2015-1076
- Drempetic, S., Klein, Ch., Zwergel B., 2019. The Influence of firm size on the ESG score: corporate sustainability ratings. Journal of Business Ethics, 167, 333-360.
- Dupuy, P., Garibal, J.-Ch., 2022. Crossdispersion biasadjusted ESG rankings. Journal of Asset Management 23 (7), 631–643.
- Kelly, B.; Pruitt, S., 2015. The three-pass regression filter: A new approach to forecasting using many predictors. Journal of Econometrics. High Dimensional Problems in Econometrics. 186 (2): 294–316.
- Kelly, B., Pruitt, S., 2013. Market Expectations in the Cross-Section of Present Values. The Journal of Finance. 68 (5), 1721–1756.
- Mann H., B., Whitney, D., R., 1947. On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat., 18 (1), 1947, 50–60.
- OECD (2022), Translating a risk-based due diligence approach into law: Background note on Regulatory Developments concerning Due Diligence for Responsible Business Conduct.
- OECD (2025), Behind ESG ratings: Unpacking sustainability metrics.
- Pedersen L., Fitzgibbons, S., Pomorski, L., 2020. Responsible investing: The ESG-efficient frontier. ttps://www.sciencedirect.com/science/article/pii/S0304405X20302853
- Ratsimiveh, K., Hubert, P., Lucas-Leclin, V., Nicolas, E., 2020. ESG scores and beyond. https://content.ftserussell.com/sites/default/files/esg_scores_and_beyond_part_1_final.pdf
- Taliento, M., Favino, C., & Netti, A. (2019). Impact of environmental, social, and governance information on economic performance: evidence of a corporate 'sustainability advantage' from europe. Sustainability, 11(6), 1738. https://doi.org/10.3390/su11061738
- Wilcoxon, F., 1945. Individual comparisons by ranking methods , Biometrics Bulletin, vol. 1, no 6, 80–83. doi 10.2307/3001968.
- Wold, S; Sjöström, M.; Eriksson, L, 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems. 58 (2), 109–130.
- Zou, H., Hastie, T., Tibshirani, R., 2006. Sparse principal component analysis. J. Com- put. Graph. Statist. 15 (2), 265–286.

Annex A. Analyzed issuers' profiles

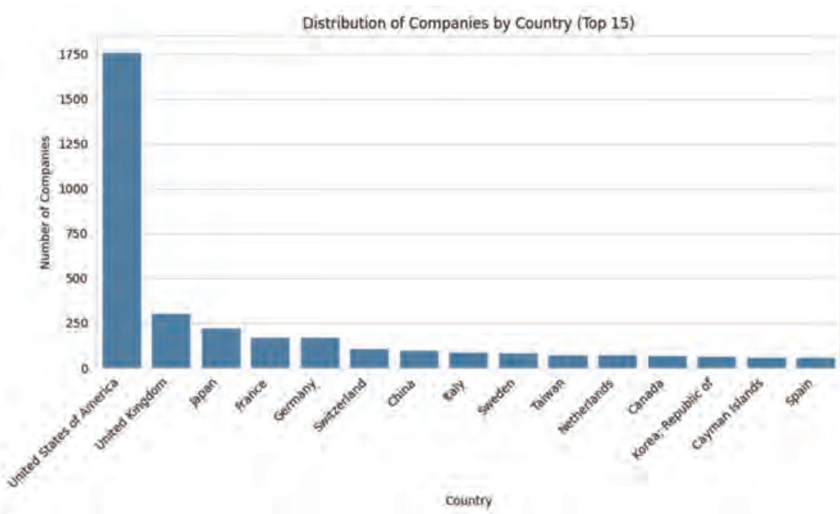
Annex A.1: Repartition by sector
Note: Activity sectors are defined using the GICS classification system.



Annex A.2: Repartition by market capitalization
Note: Nano caps are companies with capitalization < $50M, Micro caps < $300M, Small caps < $2B, Mid caps < $10B, Large caps >= $10B.



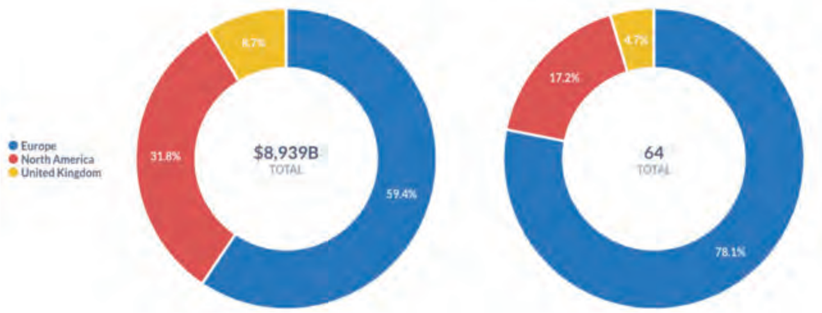
Annex A.3: Repartition by headquarter region
Note: A total of 63 countries are represented in our study dataset.



Annex B. Contribution asset managers' profiles

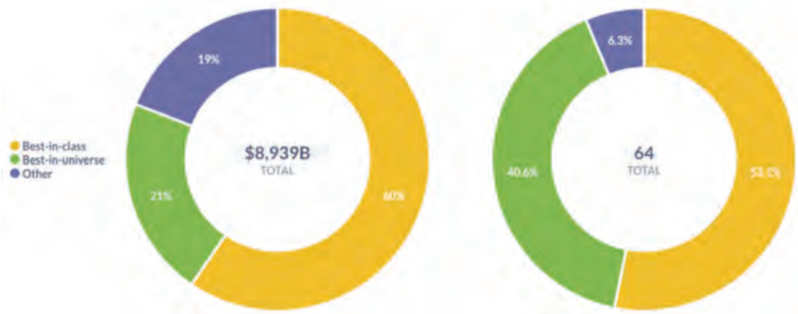
Annex B.1: Repartition by region

Note: The assets under management are determined as of 31/01/2025.



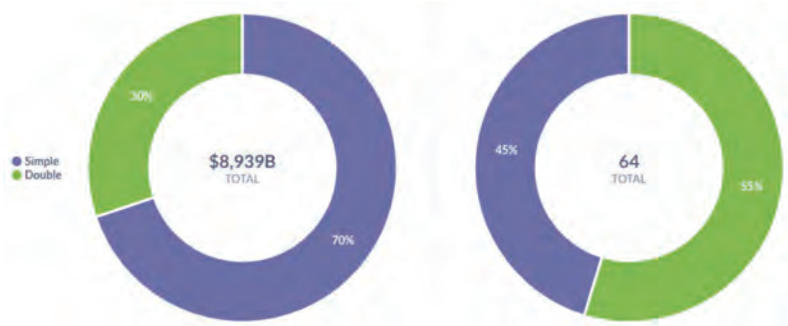
Annex B.2: Repartition by normalization strategy

Note: Best-in-class ESG ratings compare companies within the same industry or peer group, highlighting relative leaders, while best-in-universe ratings rank companies across the entire market, providing an absolute measure of performance.



Annex B.3: Repartition by materiality approach

Note: Simple materiality approaches focus on how ESG factors impact a firm's financial performance, whereas double materiality approaches also consider the broader environmental and social effects of the firm's activities.



Annex C. Correlations across proprietary ESG models

Annex C.1: Table of correlations

Note: Pairwise correlations among ESG scores from 64 proprietary ESG rating models developed by 45 asset managers. Correlations are based on the cross-sectional analysis of a set of 9089 listed companies of various countries and sectors in 2024. Min, Mean, Median and Max are respectively the minimum, mean, median and maximum levels of correlation we obtained in our sample for pairwise correlations. Standard deviation is the standard deviation of the pairwise correlations.

	Correlations
Min	-0.33
Mean	0.38
Median	0.41
Max	0.95
Standard deviation	0.26