

ESG scores and beyond (Part 1)

Factor control: Isolating specific biases in ESG ratings

July 2020

AUTHORS

Kevin Ratsimiveh

Data Scientist
+33 (0)187 448 901
kratsimiveh@lseg.com

Patrick Hubert

Senior SI Manager
+33 (0)187 448 904
phubert@lseg.com

Valéry Lucas-Leclin

Head of SI Innovation Research
+33 (0)622 027 863
vlucas-leclin@lseg.com

Emeric Nicolas

Head of SI Quantitative
Research & Analytics
+33 (0)187 448 894
enicolas@lseg.com

Overview

Following years of research and debates, there is still only an emerging consensus as to the benefits of ESG analysis, and no internationally ratified formal agreement that would stipulate what indicators should be required information for performance assessment and wider impacts evaluation.

Many studies have tried to capture its material effect on financial performance, with some success as seen in recent studies. However, before we examine materiality, it is important to ask what information ESG scores try to capture in the first place? In this study, we seek to explain what new information ESG scores bring, especially for issuers, explore how much ESG scores say about the sustainability efforts and performances of those issuers, and examine which areas of those efforts are driven by external factors that are not specific to the issuer.

The aim of these questions is not to disregard current ESG scoring and rating methodologies, but to statistically explore the role they have in broadening investors' choice and understanding.

Table of contents

Executive summary	3
What this study might bring to ESG scores	6
Exogenous drivers: a well-known bias in ESG scores that is hardly corrected	7
ESG performance assessment	10
FTSE Russell ESG Data methodology and coverage	10
Main drivers at play behind FTSE Russell ESG scores	11
Improving the relevance of ESG scores	16
Methodology	18
Results	18
Discussion	19
Residual ESG scores: a comprehensive intrinsic ESG factor	19
A combined scoring framework	23
Appendices	25
Appendix 1: an optimized Partial Least Squares (PLS) method	25
Appendix 2: a detailed study of the PLS regression outcomes	26
Appendix 3: outputs of the PLS regression's over time	29

Executive summary

Based on the FTSE Russell ESG scores, which cover nearly 4,700 companies, this report provides a) a deep analysis of the three external factors (or biases)—size, activity and country—that historically may have (strongly) influenced the ESG assessment of companies; and b) a rigorous, statistical method to analyze these factors and isolate the factor-driven component (“explained score”) and specific component (“residual score”) of any ESG score.

The ability to separate an ESG score into two components allows investors to go one step further in their analysis and decide how they want to weight or select stocks by either a) using the current ESG score that gives full credit to the company for its sustainability performance; or b) using the unbiased, adjusted score that captures only the specific efforts made by the company beyond the influence of the three main factors (size, activity, country); or c) combining the two dimensions.

For a long time, it has been commonplace to criticize ESG scores and ratings as being far from material. Many studies have tried (and still do) to capture this material effect on financial performance, with some success as seen in recent meta studies. However, before (or beyond) examining materiality, it is important to ask what information ESG scores try to capture in the first place? What new information do they bring, especially for issuers? How much do ESG scores say about the sustainability efforts and performances of those issuers? And which part of those efforts are driven by external factors that are not specific to the issuer?

ESG practitioners have found that large caps usually tend to get better ESG scores than small caps, and similarly for some sectors or subsectors. In addition, they believe that, on average, issuers from developed markets “always” scored better than those from so-called emerging markets. But are these biases justified in the first place? Where do they come from? And perhaps, more importantly for investors, could such “structural biases” prevent the selection of the most deserving issuers—those that go beyond local or regional mandatory regulations and peer-group’s common practices?

Most ESG investment solutions are still based on a best-in-class approach. If that has to be applied, one could ask whether the best-in-class issuers should not be identified after external variables like size, activity and country, have first been controlled, and controlled together, to avoid combined influence.

The purpose of these questions is not to disregard current ESG scoring and rating methodologies, but to statistically explore the role they have in broadening investors’ choice and understanding. According to our study, almost half of the information contained in the ESG ratings is explained by three factors: size, activity, and country.

Analysis of environmental, social and governance (ESG) factors has been performed for decades by—or for—investors, who thought that financial reporting alone might not provide them with enough information to assess the performance or risks of the securities they were holding. Issuers’ financial returns could suddenly be negatively affected because of overlooked ESG factors, which created, for instance, an environmental or social liability. Numerous sectors, such as in the chemical and car manufacturing sectors, the tobacco industry, pharmaceuticals, utilities etc. have experienced such actions, which over the years have raised awareness on the need to take those non-financial factors into account.

However, following years of research and debates, a) there is still only an emerging consensus as to the benefit of ESG analysis, but also b) no formal agreement has been internationally ratified that would stipulate what indicators should be required information for performance assessment and wider impacts evaluation.

This paper is the first part in a series. The other topics will address the following:

- Why and when to control ESG biases? (ESG Factor Control and Target Exposure Indexes)
- Contribution of indicators to ESG Ratings: a statistical assessment (Materiality of ESG Indicators into ESG ratings)
- Existing analysis frameworks & a proposal of improvement (ESG scores and beyond: Factor control)

Table 1. Questions addressed by this report

Questions	Answers by FTSE/BR	Rationale
Can external (exogenous) variables like size, activity, and country, play a significant role in ESG scores?	Yes	Through the statistical study of the FTSE ESG Ratings, we show that up to one half (R^2 calibration around 0.45) of the final scores can be explained by the three factors taken in combination (page 20).
Should these three variables always be controlled?	Yes, preferably but...	We think investors should know beforehand the potential effect of these three factors. However, they might be first linked to mandatory disclosures and/or peer-group best practices, so investors should not be afraid of giving credits to companies that disclose information, respect their legal requirements and, given their relative position, accept to be engaged by stakeholders. The ESG performance explained by these basic factors is still part of the overall ESG performance.
The role of each of the three specific factors is quite well known, but can they be isolated and treated separately?	Maybe	Evidence from our study shows that there is significant multicollinearity (they influence each other) between the three factors, and that a rigorous adjustment needs to take this into account. Hence the PLS regression analysis proposed in this study.
Do the three factors have a stable and persistent effect on ESG scores and ratings?	Yes	According to our study (page 29), which is based on FTSE ESG Ratings data, this effect is persistent and stable over time.
With a multi-factor (smart beta) index, size effect can be easily controlled and adjusted. Is there yet a need to control this specific size effect directly on the ESG score?	Yes and No	For an index using smart betas, the underweighting of small caps, and to a less degree mid-cap induced by ESG scorings, can be neutralized. Yet, <ul style="list-style-type: none"> a) This is not the case for active portfolios selecting a limited number of stocks (as poorly ESG-rated small caps will not make the threshold); b) Factors not taken in account, like countries for instance, still play a persistent and systemic role in the weighting; c) Because of multicollinearity, the correction brought to the size effect potentially does not correct the biases aptly in the index.
If controlled and adjusted for the three factors, do the residual ESG scores help towards a better financial performance?	Too early to answer	This question will be part of another research paper, and data can also be made publicly available for investors and academics.
If it is not possible to control for these three variables, current ESG scores could still explain Corporate Financial Performance (CFP)?	Unlikely	Now that we can show that some factors significantly influence an ESG score, establishing a direct link between ESG score and CFP logically would mean that financial performance is structurally and persistently better if companies operate as large caps, in specific activities, or under the law of some countries only.

Questions	Answers by FTSE/BR	Rationale
Is information lost once ESG scores are adjusted for the three variables?	No	<p>The initial ESG score is simply split up into two components:</p> <ul style="list-style-type: none"> • The “Explained” ESG score • The “Residual” ESG score <p>The former is explained by the three factors and can be assimilated to the level of disclosure and commitment required from the company “on average,” the latter is a specific assessment of the company’s effort.</p> <p>With this method, investors get more, no less, information, and can still recombined the two components in a final score.</p>
Is there a way to create an objective and comprehensive ESG framework that possibly avoid those three biases when assessing a company?	No, unlikely	<p>We believe that our ESG-rating framework is transparent and objectively constructed, with no specific qualitative assessment inside. However, our opinion is that at least any model based on ESG data has an implicit disclosure component (e.g. is the ESG information available?) that partly depends on requirements, either from legal regulation, stakeholders, investors, etc.</p> <p>In other words, unless generally and universally accepted ESG frameworks, like US GAAP or IFRS, are adopted globally, missing data or unveiled data may lead to a biased assessment.</p>
Can two components of ESG scores (“Explained score” & “Residual scores”) be combined without simply adding them?	Yes	<p>This option is not addressed in this report, but our opinion is that there is added value in each component as they do not tell the same story. Hence, investors can combine them in different ways, and they can probably also scrutinize the separate evolution of the two components over time.</p>
If large-cap companies on average enjoy higher ESG scores, can this explain the disconnected perception by some stakeholders of a highly rated company?	Yes, likely	<p>Our opinion is that there is a kind of “Deep Pocket” syndrome that affects large-cap companies. The bigger the issuer, the higher the responsibility, with the pressure and expectations that come along with it from all stakeholders.</p> <p>In that perspective, it seems quite logical that large caps tend to do and disclose more—if not better—to face such expectations.</p>
Is it the same to control Size and Country biases at the index building stage or at the rating stage?	Unsure. To be investigated	<p>There are sophisticated ways to control index biases, through the methodology developed by FTSE Russell known as Target Exposure Indexes.¹</p>

¹ https://content.ftserussell.com/sites/default/files/ftse_target_exposure_methodology_overview_cut_sheet_v04.pdf#page=2
<https://www.ftserussell.com/index/category/factors>.

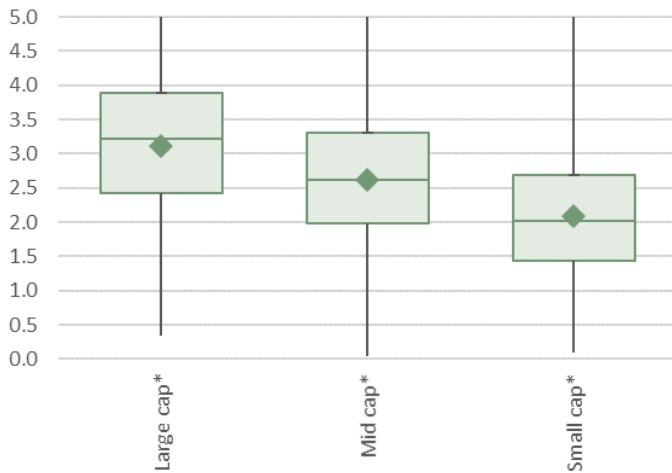
What this study might bring to ESG scores

In the following charts, we illustrate the effects of the three factors (size, activity and country) in the FTSE ESG Ratings before and the same statistics after removing the three factors.

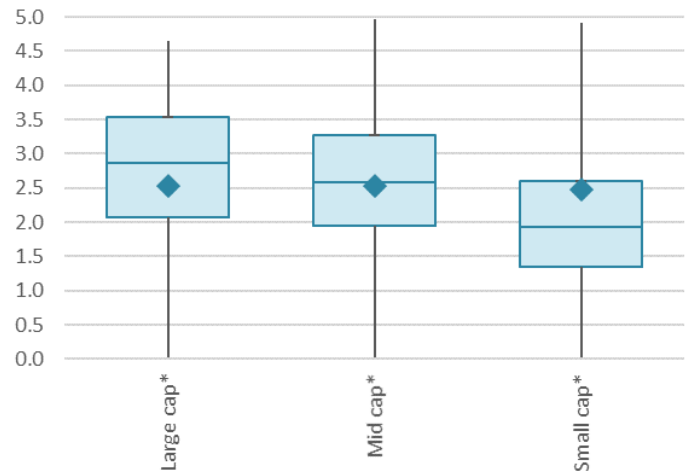
The charts below are known as box plots with whiskers, displaying logically the following features: maximum, average (diamond in boxes), average +/- 1 standard deviation, median (line in boxes), and minimum.

Charts 1 and 2. Observation of the main features of the distribution of +4,700 FTSE Russell ESG Ratings Before and After the 3-factor normalization, by size clusters*, as of June 24, 2019

Before



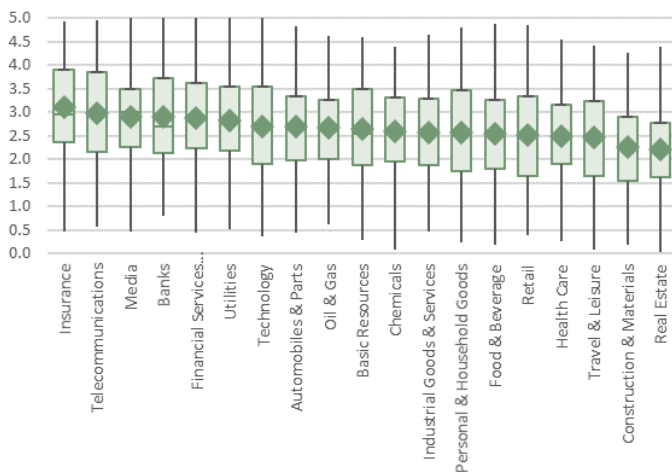
After



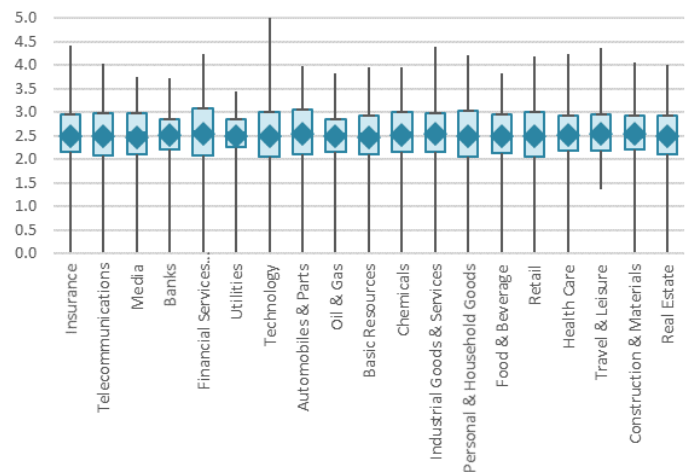
Sources: FTSE Russell, Beyond Ratings. * Companies size is based on market-cap thresholds of the companies covered by the FTSE Russell ESG database: large companies come from the 4th quartile, while small from the 1st quartile.

Charts 3 and 4. Observation of the main features of the distribution of +4,700 FTSE Russell ESG Ratings Before and After the 3-factor normalization, by ICB 2 sectors, as of June 24, 2019

Before



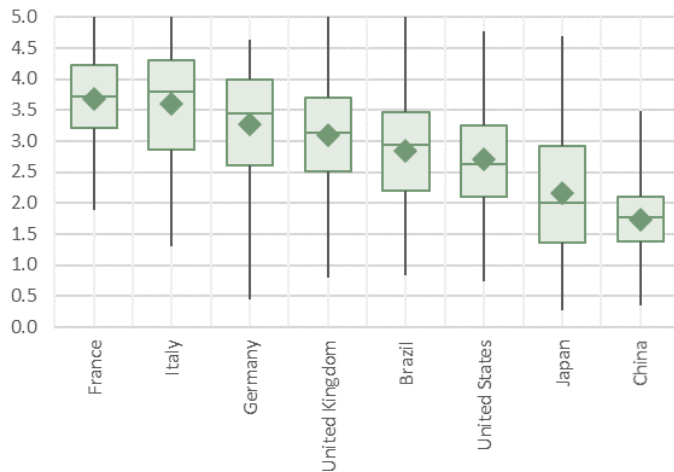
After



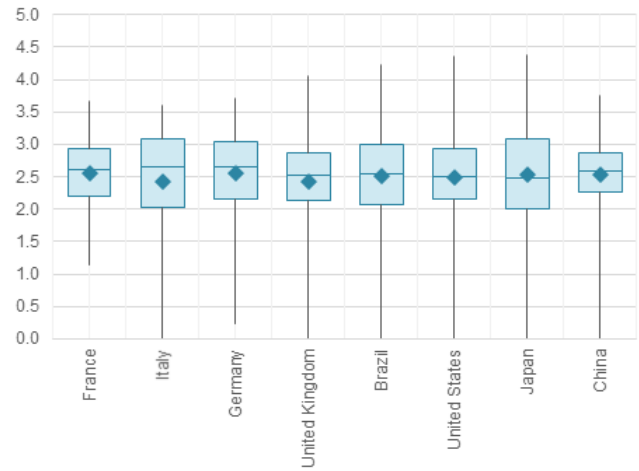
Sources: FTSE Russell, Beyond Ratings.

Charts 5 and 6. Observation of the main features of the distribution of +4,700 FTSE Russell ESG Ratings Before and After the 3-factor normalization, by countries as of June 24, 2019

Before



After



Sources: FTSE Russell, Beyond Ratings.

Exogenous drivers: a well-known bias in ESG scores that is hardly corrected

Investors and analysts have for some time underlined that ESG scores, while relevant to securities, entail various biases because of the way they are constructed. Biases can occur under various circumstances, including for instance:

- If financial accounting can sometimes be described as “aggressive” or “creative,” there are rules that set some standards when measuring financial performance. That is less true for ESG and, depending on one’s perspective, not all topics will be weighed in the same way... when they are measured;
- Issuers will report in depth and details to questionnaires received, therefore providing varying amounts of data to analyze;
- Some sectors will be more scrutinized than others in terms of their impacts and/or visibility;
- Some countries or regions have started to make reporting on various ESG issues compulsory, but not all issues and not all countries;
- When not compulsory, the various voluntary reporting frameworks might take different perspectives or stress different topics (such as human rights, climate or biodiversity), further blurring the lines for both issuers and investors or analysts.

Looking back at some investors’ or analysts’ reactions to such biases over the last few years one can find for example:

- As far back as 2006, a socially responsible investing (SRI) analysis from Société Générale, which identified, on a sample of nearly 700 European stocks, three main inherent biases in SRI ratings: 1) market capitalization, 2) country of origin and 3) sector (“SRI and valuation: the missing link?” April 2006).

- In June 2015, a paper by Credit Suisse Australia “Finding Alpha in ESG” identified an industry bias for an ESG-tilted Australian equity portfolio.²
- In September 2015, RobecoSAM announced that it had developed “an advanced ESG scoring methodology that eliminates known biases such as market cap, industry and regional biases, resulting in a new generation of unbiased ESG scores that have a low correlation to other common factors.”³
- On July 20, 2018, the Financial Times mentioned an ACCF (American Capital Formation) report outlining “inherent biases” between ratings using green criteria, highlighting company size, geographic reporting and industry sector as the three main sources of bias.^{4, 5}
- A blog by LYXOR (July 2019): SRI, data and bias: fund managers' Bermuda triangle.⁶
- Hermes (Jul 2019): Industry ESG scores for small and mid-cap companies can be misleading.⁷
- A blog by Fidelity Invest (August 2019) “ESG Ratings: Who’s driving?”⁸ stressing that ESG third-party scores should be used as enabling tools, but cannot be the overall driver for sustainable investments.
- More recently two important studies have outlined the extent to which some of the well-known ESG ratings available on the market strongly disagreed together, with a level of correlation surprisingly low, at only 42.9% for six rating providers on the S&P500 index over 2013-2017 (Krueger, Riand, Schmidt, 2020).⁹ An influence from the country factor was put forward, both for corporates and the raters. Similar conclusions were delivered by the MIT (Berg, Koelbel, and Rigobon, 2019): the correlations between the ratings are on average 0.54 and range from 0.38 to 0.71 among the six rating agencies selected.¹⁰ Yet the authors have preferred investigating the influence of the “raters” (scope, weighting, universe), before daring to control for biases.

² https://researchplus.credit-suisse.com/rpc4/docView?language=ENG&format=PDF&document_id=1049893651&source_id=emcms&serialid=EH1lrEKQ2OShF3%2bmR54mSQR%2frlqHN7EaFHalvpgxxtE%3d

³ <https://www.robecosam.com/en/about-us/history.html>

⁴ <https://www.ft.com/content/a5e02050-8ac6-11e8-bf9e-8771d5404543>

⁵ http://accfcorgov.org/wp-content/uploads/2018/07/ACCF_RatingsESGReport.pdf

⁶ <https://www.securities-services.societegenerale.com/en/insights/views/news/sri-data-and-bias-fund-managere-bermuda-triangle/>

⁷ <https://www.corporatesecretary.com/articles/esg/31709/hermes-industry-esg-scores-small-and-mid-cap-companies-can-be-misleading>

⁸ <https://www.fidelityinternational.com/blog/esgenius-ratings-whos-driving-d1679c-en5/>

⁹ Gibson, Rajna and Krueger, Philipp and Riand, Nadine and Schmidt, Peter Steffen, ESG Rating Disagreement and Stock Returns (December 22, 2019). Swiss Finance Institute Research Paper No. 19-67; European Corporate Governance Institute – Finance Working Paper No. 651/2020. Available at SSRN: <https://ssrn.com/abstract=3433728> or <http://dx.doi.org/10.2139/ssrn.3433728>.

¹⁰ Berg, Florian and Kölbel, Julian and Rigobon, Roberto, Aggregate Confusion: The Divergence of ESG Ratings (May 17, 2020). Available at SSRN: <https://ssrn.com/abstract=3438533> or <http://dx.doi.org/10.2139/ssrn.3438533>.

Regulators too, have expressed reservations on the accuracy of ESG ratings/scores. On May 28, 2020, Jay Clayton, Chairman of the US Securities and Exchange Commission, was thus quoted by the Financial Times: *“I have not seen circumstances where combining an analysis of E, S and G together, across a broad range of companies, for example with a ‘rating’ or ‘score’, particularly a single rating or score, would facilitate meaningful investment analysis that was not significantly over-inclusive and imprecise.”*

Intuitively, one can surmise what the rationale behind each of the three main factors (size, country, sector) might be:

- **Size:** larger organizations tend to have better reporting and communication structures. Besides, the more a company is exposed to the market, the more its shareholders and stakeholders are likely to scrutinize its activities and seek information. That incentive to report is compounded by the fact that more analysts follow larger issuers, providing more external opinions and data than for smaller ones: most companies will prefer to provide their own factual evidence rather than see ill-founded hypotheses damage their score.
- **Country:** corporate culture, governance expectations and regulatory requirements are not at the same level between countries. Some European countries for instance require companies that are listed on their markets to report key ESG metrics (e.g. Article 273 in France on GHG emissions), whereas other markets authorities have not implemented such requirements. The level of economic and social development will also influence a company’s underlying ESG scores based on its headquarters’ location, when factors like health, education, working conditions, tax fairness, human rights etc. are considered. Another factor differentiating nations is the sectoral exposure of a country: an issuer from commodity-exporting Australia, for instance, will not be perceived in the same way as its competitor from import-dependent Japan.
- **Sector:** when it comes to ESG, some activities are more exposed—favorably or unfavorably—to ESG issues than others and thus will either benefit from a supposedly positive view (e.g. solar energy or water utilities) or could tend to report more to compensate an image deficit (e.g. oil & gas, mining or gambling).

Surprisingly, perhaps given the rising importance of ESG as an investment theme, so far there seems to be scant academic evidence (apart from above-mentioned studies) of how much size, country and sector account for in an issuer’s overall score. Using FTSE Russell’s ESG data for listed companies as the primary material for this paper, we focused on those three factors (individually and combined) to ascertain their influence in the ESG scores and, ultimately, determine an **intrinsic ESG score** for each company that would remove such influence.

(We should point out that our approach is purely concentrated on the ESG scores and the overall influence that size, country and sector exert on them. At this stage its perspective is different from that of FTSE Russell’s research on **Target Exposure Factor** and its application to create a set of index tools which, using tilts, design benchmarks to maintain a constant level of targeted active factor exposure against a given index, while minimizing off-target consequential exposures. The factors targeted can be Quality, Value, Momentum, Low Volatility, Size... or ESG, and the overall objective is performance for investors, giving priority to those set factors. On the other hand, this paper only deals with ESG scoring and potential ways to improve its relevance.)

ESG performance assessment

FTSE Russell ESG Data methodology and coverage

FTSE Russell has been analyzing ESG performance for over 15 years; its ESG Ratings can be accessed through an online data model.¹¹

The database comprises over 7,200 securities in 47 developed and emerging markets, represented by constituents from the FTSE All-World, FTSE All-Share and Russell 1000® Indexes.

A FTSE Russell **ESG rating** provides an objective measure in a single figure, which is the cumulative calculation of an issuer's total ESG exposure and performance in multiple dimensions. That figure rests on assessments conducted for three pillars, covering 14 themes, spread over 300 indicators.

Three Pillars (E, S and G)

For each of the Environmental, Social and Governance pillars, two figures are produced:

- **Exposure**, which measures the relevance of the Pillar issues for a company (from 0 = none, to 3 = high)
- **Score**, which measures the quality of a company's management of the Pillar issues (from 0 = no disclosure, to 5 = best practice)

Industry relative ESG Ratings and Scores: In addition to the “absolute” Scores and Ratings described above, peer relative Scores and ESG Ratings are also calculated by comparing a company's Score or ESG Rating to others within the same FTSE Industry Classification Benchmark (ICB) Supersector.¹² The overall ESG Rating is represented as a percentile where a “1” indicates that a company is in the bottom 1% and “100” indicates a company is in the top 1%.

14 Themes¹³

In turn, each pillar aggregates several of the 14 themes pertinent to it for which, in the same way, an **Exposure** and a **Score** measure have been produced.

300+ indicators

The Pillars and Themes are themselves built on over 300 individual indicator assessments, that are applied to each company's unique circumstances. Each Theme contains 10 to 35 indicators, with an average of 125 indicators being applied per company.

The underlying ESG Ratings and data model allows investors to understand a company's exposure to, and management of, ESG issues in multiple dimensions.

The model also enables “slice and dice” to meet each user's needs to extract ESG data at multiple levels to assess and apply it in a variety of ways (e.g. Portfolio evaluation and manager

¹¹ <https://www.ftserussell.com/data/sustainability-and-esg-data>.

¹² The ICB offers a framework for comparing and analyzing like-organizations, at various levels (11 Industries, 20 Supersectors, 45 Sectors and 173 Subsectors).

¹³ ESG themes are (in alphabetical order): Anti-Corruption, Biodiversity, Climate Change, Corporate Governance, Customer Responsibility, Environmental Supply Chain, Health & Safety, Human Rights & Community, Labour Standards, Pollution & Resources, Risk Management, Social Supply Chain, Tax Transparency, Water Use.

due diligence, Engagement and stewardship, Risk Management, Research and analysis, Active portfolio management, Custom benchmarks etc.).

See FTSE Russell ESG Data Model 6th Research Cycle (2019/20)¹⁴ and update¹⁵ for a more detailed overview of the methodology.

Main drivers at play behind FTSE Russell ESG scores

In this section, we analyze the influence of each identified factor—the size, geographical location and sector of an issuer—in the level of its ESG overall score based on the FTSE Russell ESG database.

We carried out our study at the end of the first half of 2019 on nearly 4,700 securities (4,684 share listings on 50 stock exchanges throughout the world).

Size influence

We measure the size of a corporate issuer by its market capitalization. As the market cap units vary by a large factor (10^5 , i.e. from millions to hundreds of billions), we smooth its distribution by using the natural logarithm of the market cap for each issuer in our panel as the size variable.

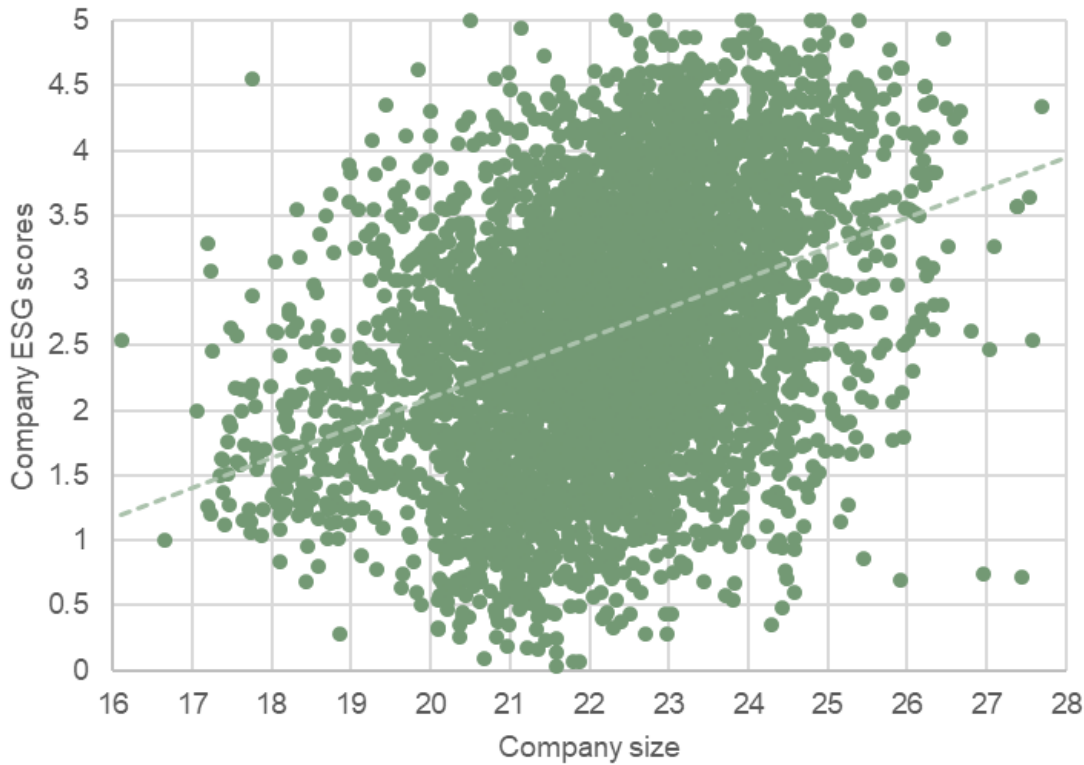
The scatter plot below (Chart 7) shows the relation between issuers' ESG scores and issuers' size, based on the companies covered by the FTSE Russell ESG database.

The link between the two variables is outlined by the correlation coefficients (Table 2) associated to the relation. Furthermore, the r-squared associated to the relation sets at 0.15—that can be interpreted as overall 15% of the variance in issuers' ESG scores—can be explained by their size.

¹⁴ <https://qsd.ftserussell.com/Docs/ESG/FTSE%20Russell%20ESG%20Data%20Model%20Methodology%20-%20April-2019-March-2020-Rev2.pdf>.

¹⁵ https://qsd.ftserussell.com/Docs/ESG/FTSE%20Russell%20ESG%20Ratings_Methodology%20changes%20summary%202019-2020.pdf.

Chart 7. Company ESG scores and company size measured by the natural logarithm of market cap. The line represents the regression line, data as of June 24, 2019



Sources: FTSE Russell, Beyond Ratings.

Table 2. Correlation coefficient between company ESG scores and company size measured by the natural logarithm of market cap, data as of June 24, 2019

Type	Value
Pearson ¹⁶ correlation	0.39
R-Squared ¹⁷	0.15

Sources: FTSE Russell, Beyond Ratings.

¹⁶ In statistics, the Pearson's correlation coefficient measures the linear relationship between two variables. It has a value between +1 to -1, where 1 is a total positive linear correlation, -1 a total negative linear correlation and 0 no linear correlation. The larger the absolute value of the coefficient, the stronger the relationship.

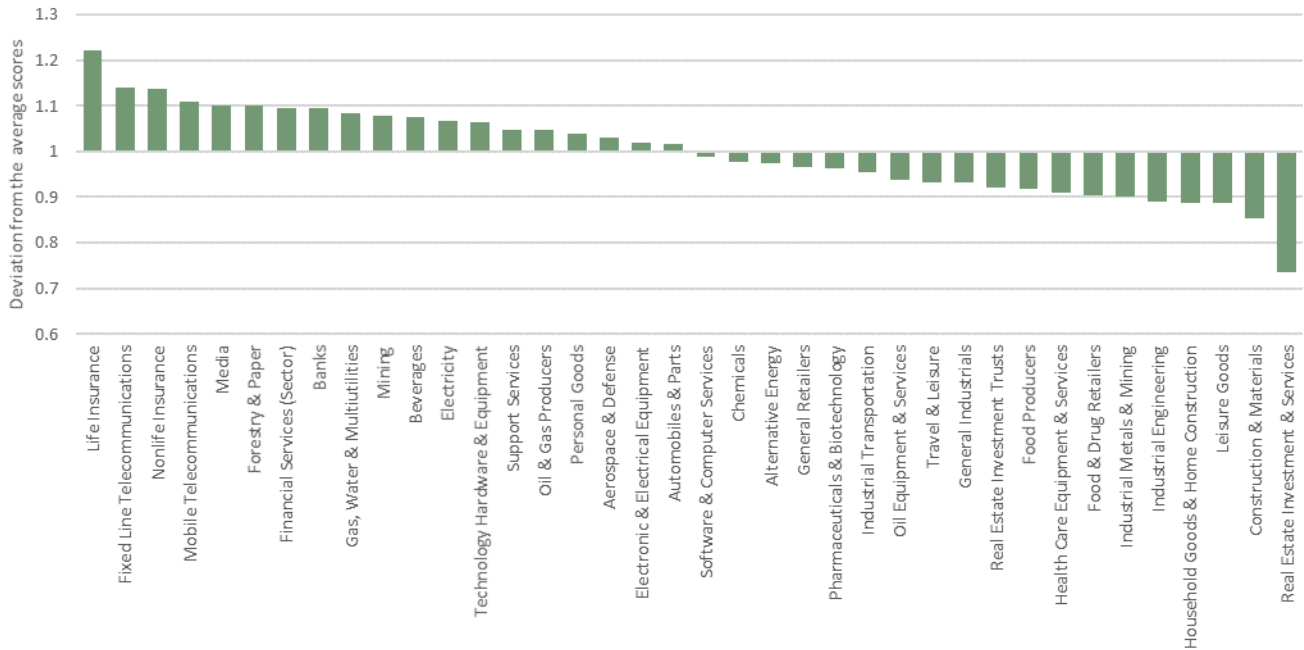
¹⁷ In statistics, the coefficient of determination or "R-squared", measures how close the data are to the fitted regression line.

Sector effect

To measure the influence of the sector on the level of ESG scores, we use the ICB classification at level 3, which distinguishes 39 different sectors.¹⁸

We compute average ESG scores by sector and compare them through the distribution chart below (Chart 8). It shows sectors with the highest to lowest averages (from left to right). To avoid a misleading comparison, we consider the average of the sector if, and only if, the sector comprises at least 20 issuers.

Chart 8. Distribution of ESG scores by ICB level 3 sectors: deviation from the average as of June 24, 2019

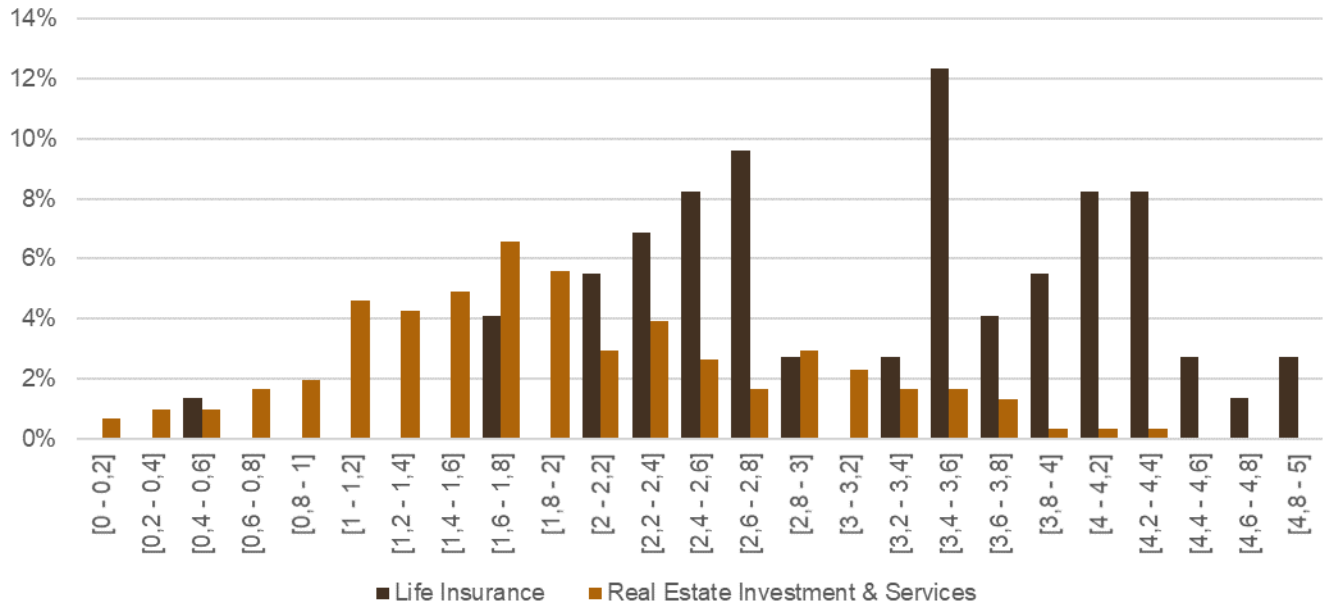


Sources: FTSE Russell, Beyond Ratings.

The chart displays large disparities between sectors. For example, Chart 9 compares the distribution of ESG scores between the sector with the highest average ESG score to the one with the lowest average ESG score. As can be seen, the major part of Life Insurance companies in our panel has scores above 2.2, while most companies in the Real Estate Investment and Services sector have scores below 2.4.

¹⁸ Our study used the then current ICB taxonomy which has since been revised (now 45 sectors).

Chart 9. Distribution of ESG scores for Life Insurance and Real Estate Investment & Services companies, (data as of June 24, 2019)



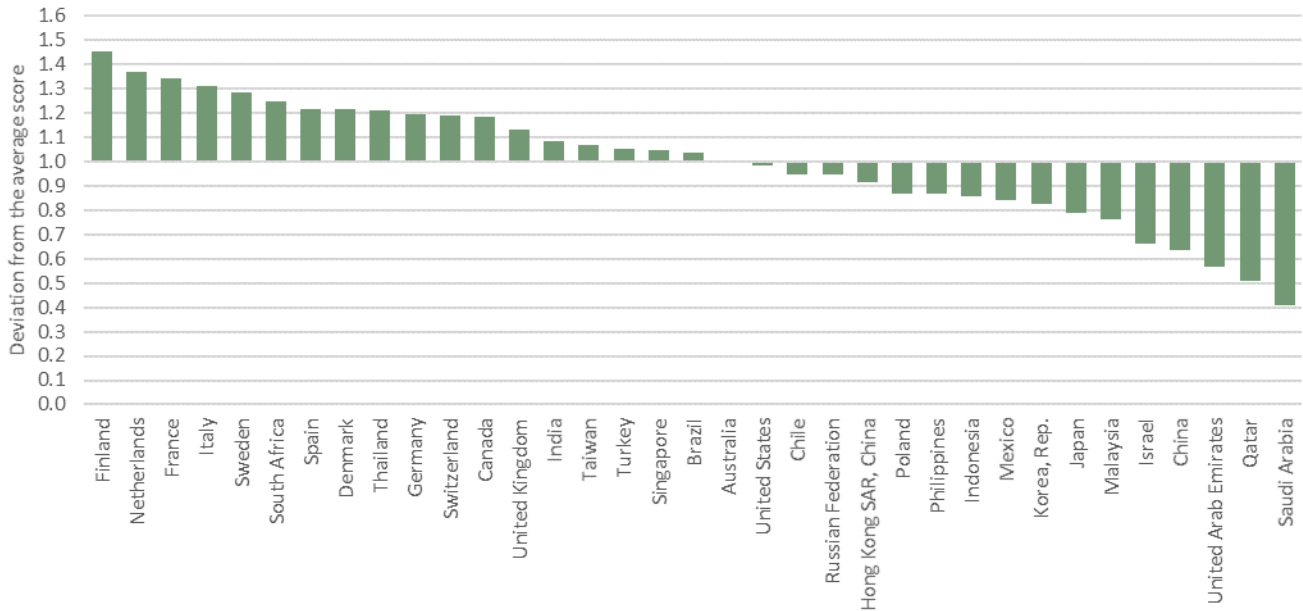
Sources: FTSE Russell, Beyond Ratings.

Geographical effect

The location of a company impacts its ESG behavior, reporting and, ultimately, performance. To test this hypothesis, we use the main country of listing of each issuer covered by the FTSE Russell ESG database as geographical variable.

In the same way as for the sector analysis, we compute average ESG scores by country if, and only if, it comprises at least 20 issuers and compare them through a distribution chart (Chart 10).

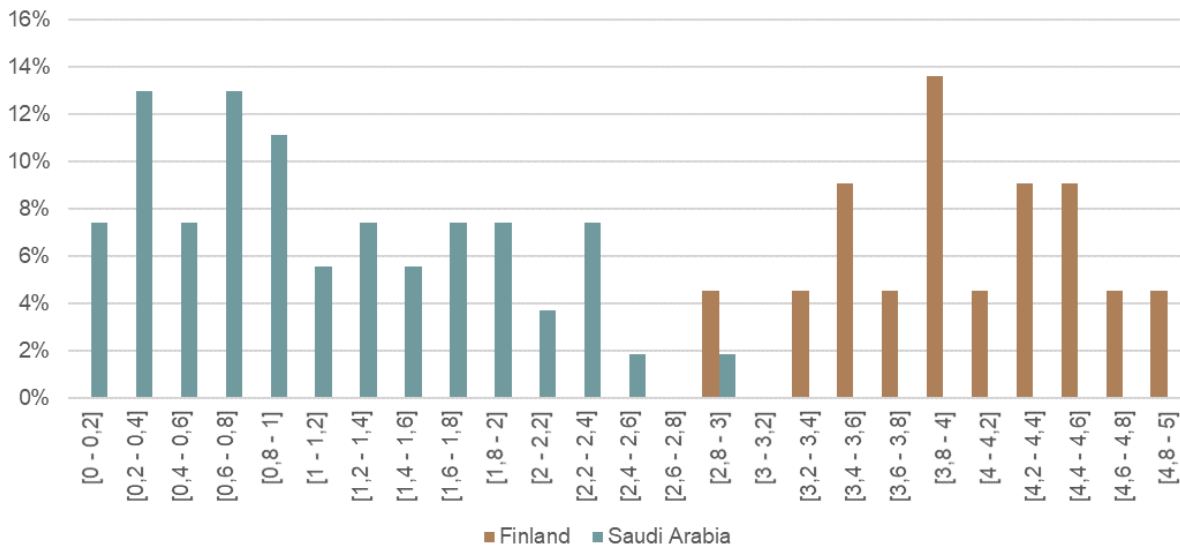
Chart 10. Distribution of ESG scores by countries: deviation from the average as of June 24, 2019



Sources: FTSE Russell, Beyond Ratings.

The chart shows large differences between countries. This is confirmed by Chart 11 which shows the distribution of ESG scores between the country with the highest average ESG score to the one with the lowest average ESG score. Most companies based in Finland have scores above 2.4, while the majority of companies based in Saudi Arabia have scores below 2.4.

Chart 11. Distribution of ESG scores for Finland and Saudi Arabia companies as of June 24, 2019



Sources: FTSE Russell, Beyond Ratings.

Improving the relevance of ESG scores

As shown by the literature and confirmed by tests performed on the FTSE Russell ESG dataset, size, sector and geographical location constitute the main drivers of ESG scores in that they highly influence the ESG performance of a corporate issuer. We propose here a statistical method to systematically extracting the part of ESG scores that are not linked to these factors.

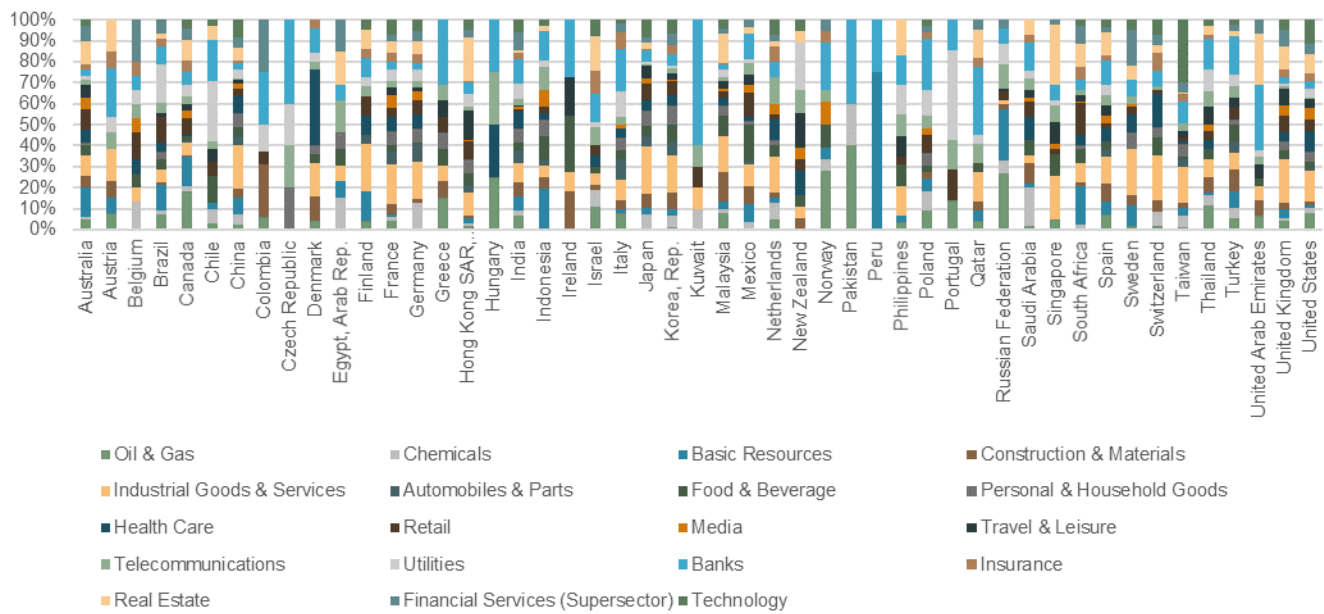
A common way to remove such “biases” is to perform a regression analysis with the level of individual ESG scores as response variable and the three factors as regressors. Indeed, the goal of regression analysis is to isolate the relationship between each independent variable (regressor) and the dependant one.

A key hypothesis in this kind of analysis is the independence of the regressors: they must be uncorrelated otherwise a multicollinearity¹⁹ issue could occur.

However, size, sector and geographical factors seem to be correlated in that companies’ size depend both on geographical location and sectors.

Indeed, sectors’ spread is not the same across countries (Chart 12).

Chart 12. Weight* of each ICB (level 2) sector by countries, as of June 24, 2019

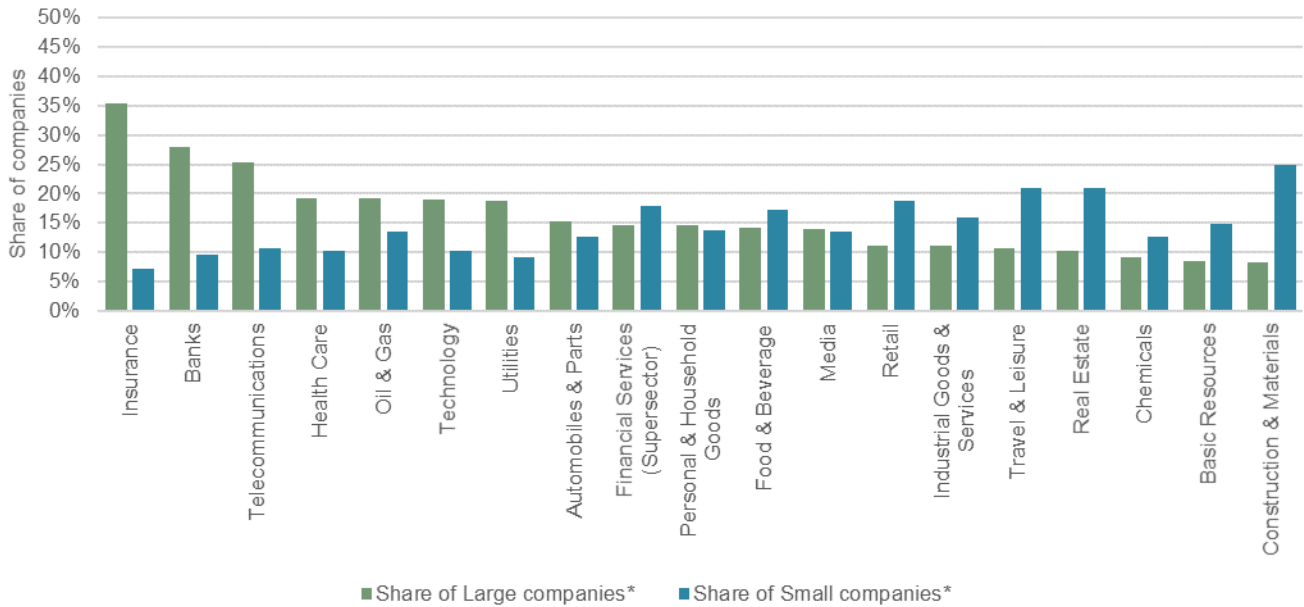


Sources: FTSE Russell, Beyond Ratings. * Weights are based on the number of companies in each sector.

Furthermore, large companies tend to come from financial, Oil & Gas and IT sectors, while small ones are more related to industrial sectors (Chart 13).

¹⁹ Multicollinearity occurs when a model includes multiple factors that are correlated not just to the dependant variable but also to each other: it results when (one or more) factors are redundant to some extent.

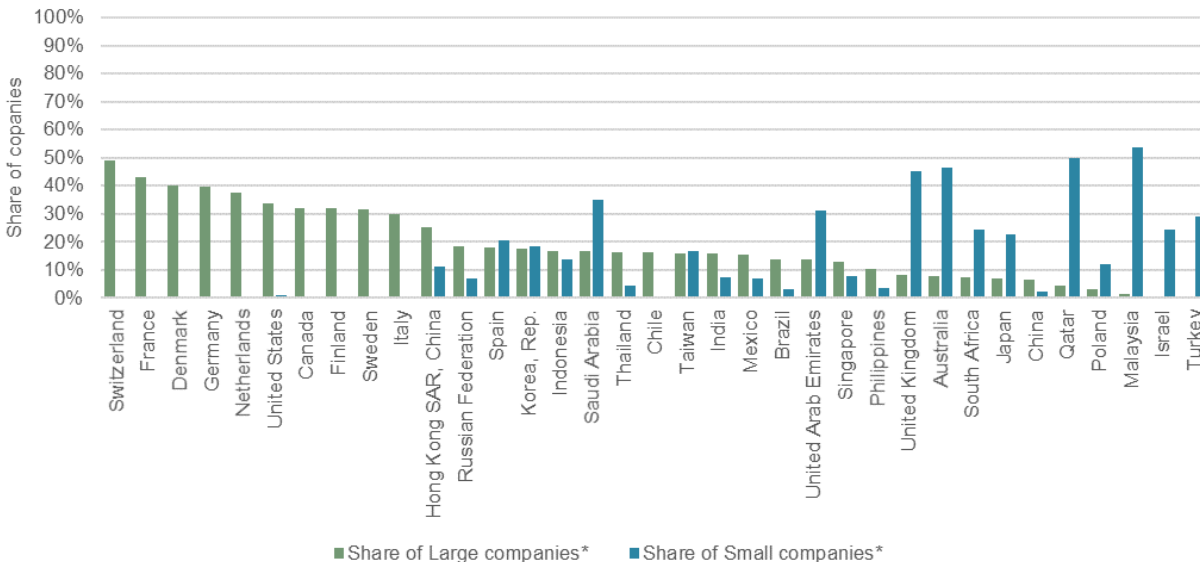
Chart 13. Share of large* and small* companies by ICB 2 sectors as of June 24, 2019



Sources: FTSE Russell, Beyond Ratings. * Companies size is based on market-cap thresholds of the companies covered by the FTSE Russell ESG database: large companies come from the 4th quartile, while small from the 1st quartile.

Finally, large companies are mostly based in developed markets and small companies in emerging market countries (Chart 14).

Chart 14. Share of large* and small* companies by countries as of June 24, 2019



Sources: FTSE Russell, Beyond Ratings. * Companies size is based on market cap thresholds of the companies covered by the FTSE Russell ESG database: large companies come from the 4th quartile, while small from the 1st quartile.

Methodology

To deal with this apparent multicollinearity and to discard redundant regressors, we propose an optimized Partial Least Squares (PLS) regression method.

PLS regression is an alternative statistical method to Ordinary Least Squares (OLS) that fits a multiple linear regression model on a set of relevant independent regressors.

It could be viewed as a two-steps process that:

- First, forces the set of initial regressors to be independent by aggregating them into relevant components and;
- Then, models the relationship between the response variable and the set of “new” regressors.

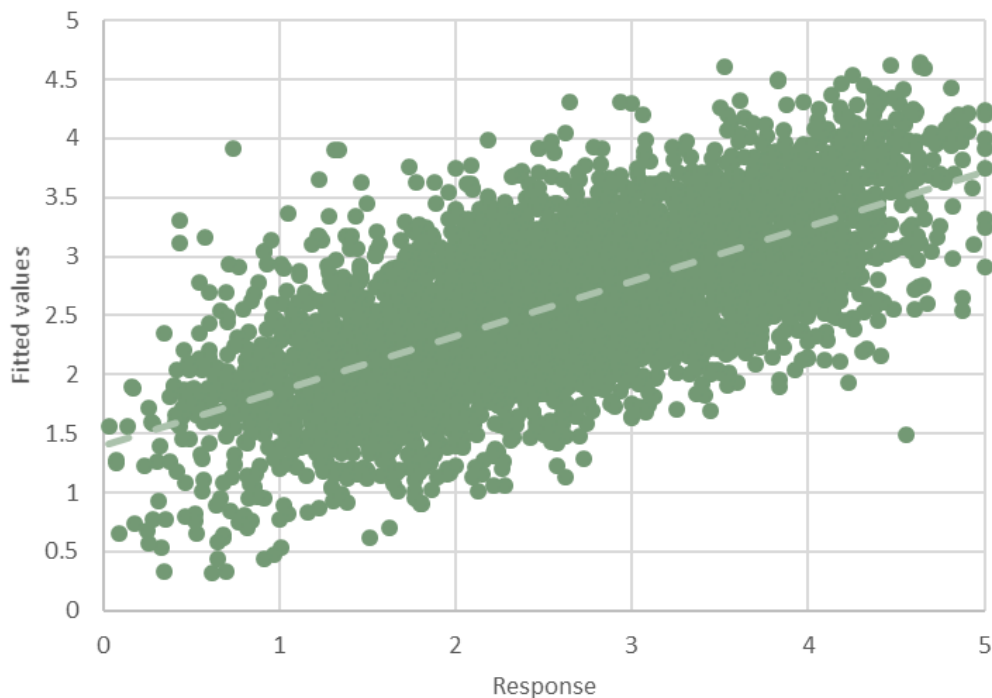
Using machine learning techniques, we propose to optimize the calibration of the PLS regression.

The aim of the algorithm is to minimize the cross-validation Mean Squared Error²⁰ (MSE) by varying simultaneously the number of components and the set of initial regressors. Indeed, for each combination, we add a cross-validation step by testing the predictive ability of the calibration on random sub-samples and keeping the one with the minimum cross-validation MSE. See Appendix 1 for a more detailed overview of the methodology.

Results

We compute fitted values from the PLS regression using the optimal number of components and combination of regressors found by the algorithm. These fitted values constitute the part of ESG scores that can be attributed to the size, sector and geographical factors. Chart 15 plots in variable against fitted values. It shows the predictive ability of the calibration model in that we could deduce a clear path from the relationship.

Chart 15. Response variable and fitted values from the PLS regression



Sources: FTSE Russell, Beyond Ratings.

²⁰ The mean squared error (MSE) of a model is the average deviation between the response variable and the fitted values.

In addition, to confirm the quality of our model, we compute the coefficient of determination (r-squared) and the mean squared error (MSE) between response variable and fitted values.

As we added a cross-validation step, we also get cross-validation fitted values, noise and thus cross-validation quality metrics.

The r-squared associated with our model sets at 0.46. This can be interpreted as 46% of variance in companies' ESG scores being explained by their size, sectors and geographical location. Thus, by adding sectors and geographical location, the predictive ability of the model improves by c. 0.4 compared to the model with only companies' size as a regressor.

Furthermore, both MSE in calibration and cross-validation are especially low with regards to the scale of the ESG scores.

Table 3 below summarizes quality metrics from our model.

Table 3. Quality metrics from the PLS regression

Type	Value
R-Squared <i>calibration</i>	0.46
R-Squared <i>cross-validation</i>	0.44
MSE <i>calibration</i>	0.54
MSE <i>cross-validation</i>	0.56

Sources: FTSE Russell, Beyond Ratings.

To go further than the model's [overall] performance analysis, we study in Appendix 2 the importance of independent variables in the model and their estimated link with the response variable.

Furthermore, given the availability of historical series within the FTSE Russell ESG database, we study the stability of the PLS regression's outputs over time in Appendix 3. It shows that both the quality of the regression and the structure of the coefficients are relatively stable over time.

Discussion

Based on the panel of companies covered by the FTSE Russell ESG database and through an optimized PLS regression method, we can distinguish the part of ESG scores explained by size, sector and geographical factors from the "idiosyncratic" part.

Residual ESG scores: a comprehensive intrinsic ESG factor

We propose to transform residuals from the PLS regression to extract a comprehensive intrinsic factor; the resulting measure arguably represents an issuer's **Residual ESG** score.

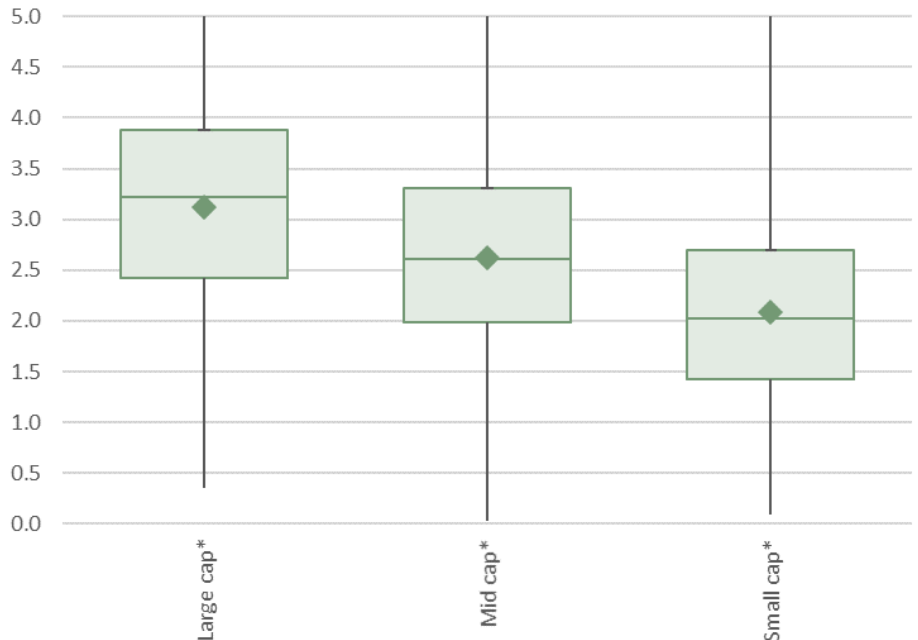
To set them between 0 to 5 like the scale of both initial ESG scores and their estimates, we define an issuer's **Residual ESG score** as the linear transformation of its residuals (E):

$$\text{Residual ESG Score}_i = 5 \times \left(\frac{E_i - \max(E)}{\max(E) - \min(E)} \right)$$

This specification has large impact on the distribution of the ESG performance across sector, geographical location and size level, as shown by the box plots (Charts 16 to 21) below.

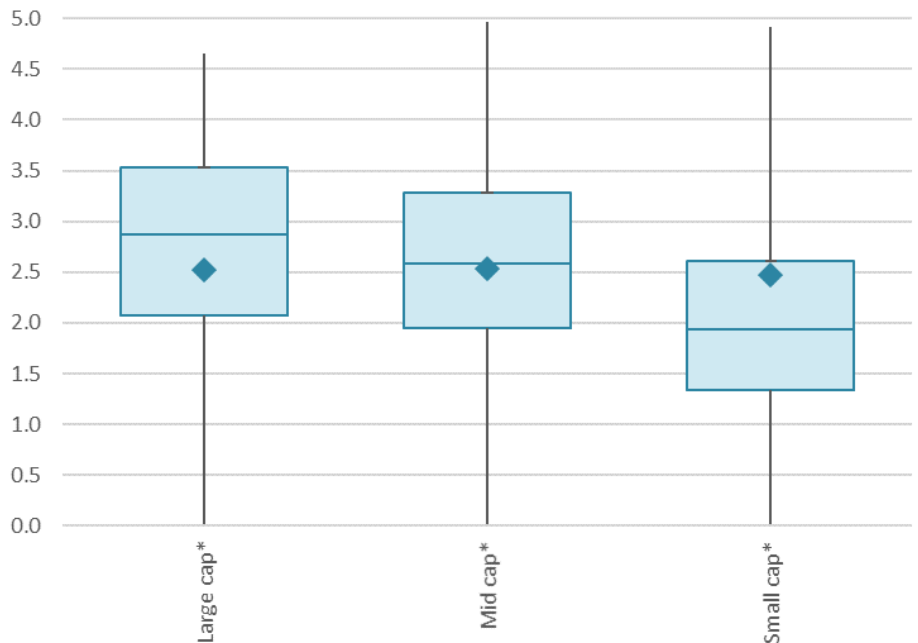
The whiskers of these box plots display the following features of initial and residual ESG scores: maximum, average (diamond shape in the box), median and minimum.

Chart 16. Overview of initial ESG scores by size clusters* as of June 24, 2019



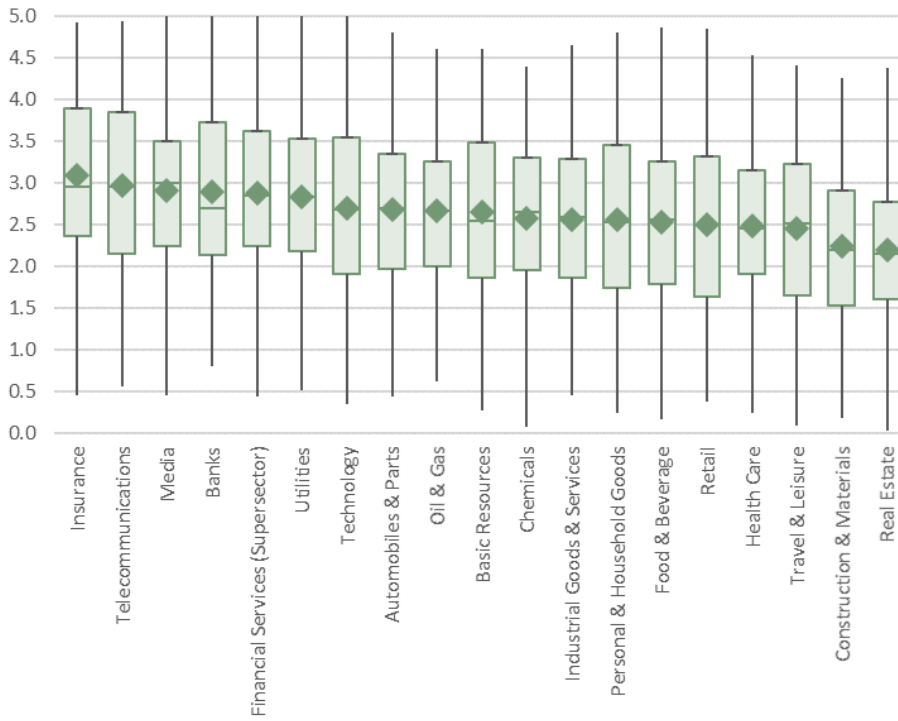
Sources: FTSE Russell, Beyond Ratings. * Companies size is based on market-cap thresholds of the companies covered by the FTSE Russell ESG database: large companies come from the 4th quartile, while small from the 1st quartile.

Chart 17. Overview of Residual ESG scores by size clusters* as of June 24, 2019



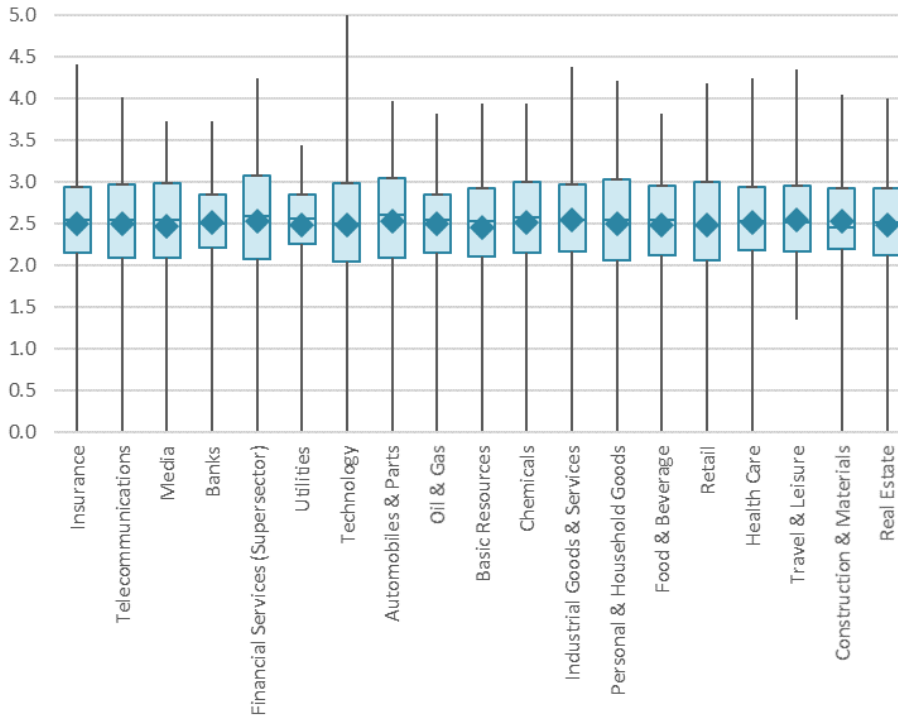
Sources: FTSE Russell, Beyond Ratings. * Companies size is based on market-cap thresholds of the companies covered by the FTSE Russell ESG database: large companies come from the 4th quartile, while small from the 1st quartile.

Chart 18. Overview of initial ESG scores by ICB (level 2) sectors as of June 24, 2019



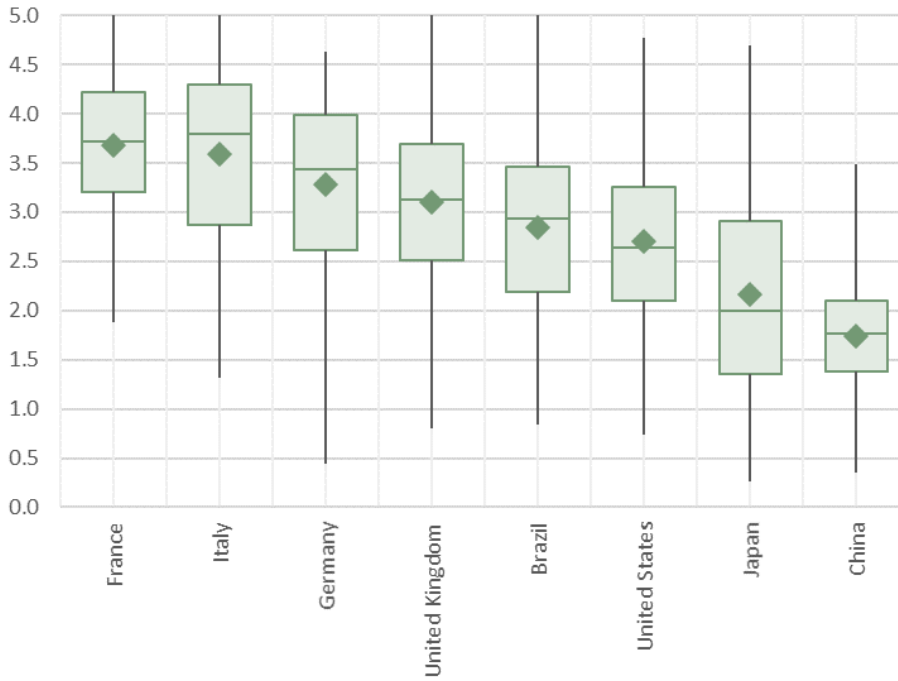
Sources: FTSE Russell, Beyond Ratings.

Chart 19. Overview of Residual ESG scores by ICB (level 2) sectors as of June 24, 2019



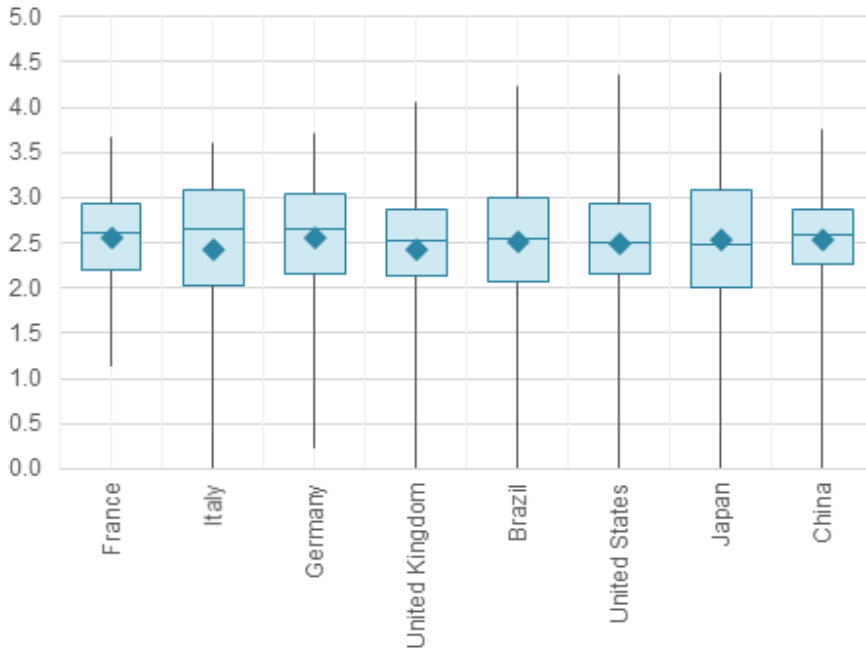
Sources: FTSE Russell, Beyond Ratings.

Chart 20. Overview of initial ESG scores by countries as of June 24, 2019



Sources: FTSE Russell, Beyond Ratings.

Chart 21. Overview of Residual ESG scores by countries as of June 24, 2019



Sources: FTSE Russell, Beyond Ratings.

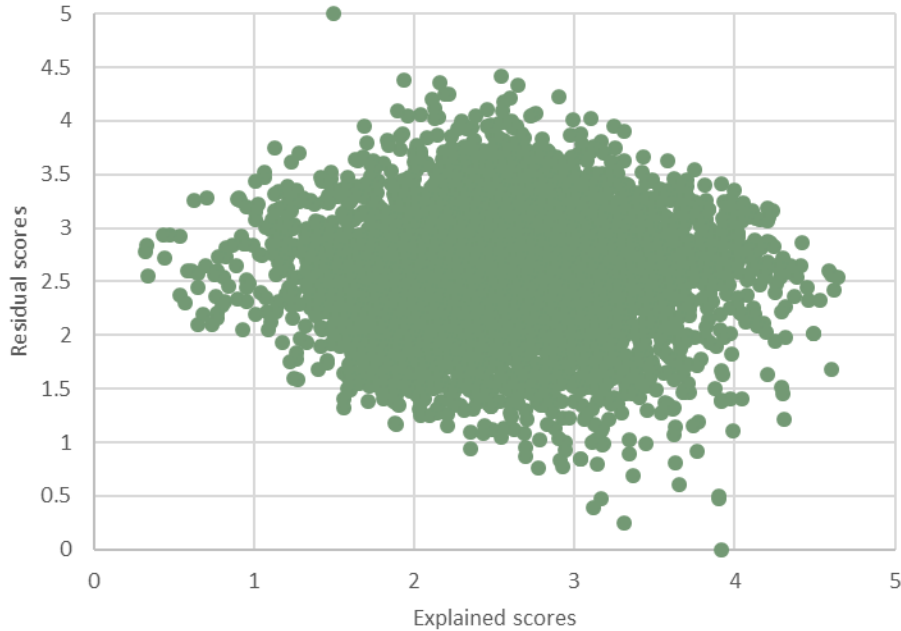
A combined scoring framework

Residual ESG scores can be combined with the fitted values from the PLS algorithm to improve the relevance of companies' ESG performance assessments.

Indeed, while Residual ESG scores represent the ESG performance of each company independently of the three factors, the fitted values estimate what each company's ESG performance would be if only the three factors were considered. We propose to define these latter as **Explained ESG scores**.

Chart 22 plots the Explained ESG scores against Residual ones.

Chart 22. Explained scores vs. Residual score



Sources: FTSE Russell, Beyond Ratings.

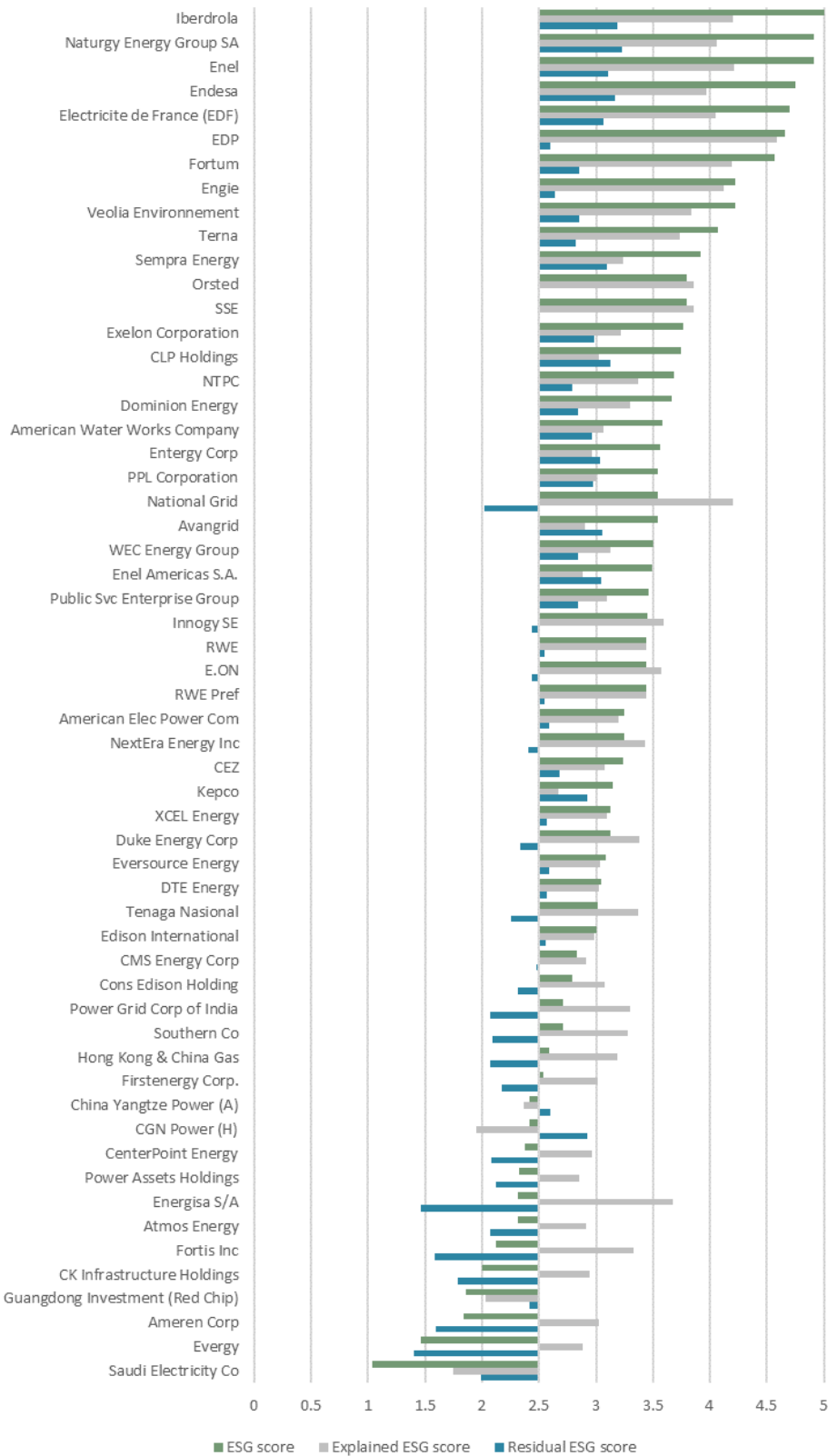
Thanks to this framework, we can highlight companies that perform well from an ESG perspective with respect to the performance of their sector, geography and size:

- In the upper-right side: companies that are part of sectors, located in countries and of a size that would allow them to have a good ESG performance and that have a good intrinsic performance;
- In the lower-right side: companies that should perform well with respect to their sector, geographical location and size but have weak intrinsic performance;
- In the upper-left side: companies whose weak Explained ESG performance should be nuanced by size, sector and geographical factors;
- In the lower-left side: companies that have both weak Explained and Residual scores.

Furthermore, Residual and Explained ESG scores show great stability over time, comparable to the level of stability of the initial ESG scores (see Appendix 3).

Chart 23 provides a focus on large cap companies from the Utilities industry covered by the FTSE Russell ESG database.

Chart 23. ESG scores, explained scores and residual scores for large cap* companies from Utilities industry covered by the FTSE Russell ESG database, as of June 24, 2019



Sources: FTSE Russell, Beyond Ratings. * Companies size is based on market cap thresholds of the companies covered by the FTSE Russell ESG database: large companies come from the 4th quartile while small from the 1st quartile.

Appendices

Appendix 1: an optimized Partial Least Squares (PLS) method

Multiple linear regression methods aim at building a linear model: $Y = X\beta + \varepsilon$.

Using matrix representation, X is a matrix gathering all predictors vectors, Y the response vector, β the vector of estimated parameters and ε the noise vector.

The quality of a calibration model is then measured using the coefficient of determination (R^2) and the mean squared error (MSE) which is the average deviation between the response variable and the fitted values, estimated by the model: $MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ where $\hat{Y}_i = \beta X_i$ and n the sample size.

In this study, the set of predictors vectors contained in the matrix X are highly correlated and may obscure the relation we want to highlight (multicollinearity issue).

To handle this multicollinearity issue, PLS regression transforms the initial set of predictors X into an equivalent set $X' = XW$ through a linear transformation W , such that the ultimate set of predictors X' (which are the principal components) are linearly independent. The linear transformation W is found by maximizing the covariance between the response variable Y and the principal components X' .

However, the main risk that could skew results from a PLS model is overfitting. In machine learning, an overfitting issue occurs when the model describes too well the response variable in that the model contains more parameters than can be justified by the data. In such case, the model has extracted some of the noise by considering it as an underlying structure.

A common procedure to prevent an overfitting issue is to discard initial predictors that do not give more information and thus are redundant to some extent.

To manage both multicollinearity and overfitting issues, we build an algorithm that optimizes the calibration of the PLS regression. It aims at finding the optimal number of components given an optimal set of initial regressors that minimizes the cross-validation MSE.

More precisely, the cross-validation procedure consists of splitting the initial sample of data into few random sub-samples, leaving one of the sub-samples out and fitting a model on the remaining sub-samples. The model is then used to predict the values of the left-out sub-sample. The process is repeated²¹ so that all samples have been predicted once. Finally, the cross-validation MSE is the average MSE across all tested sub-samples.

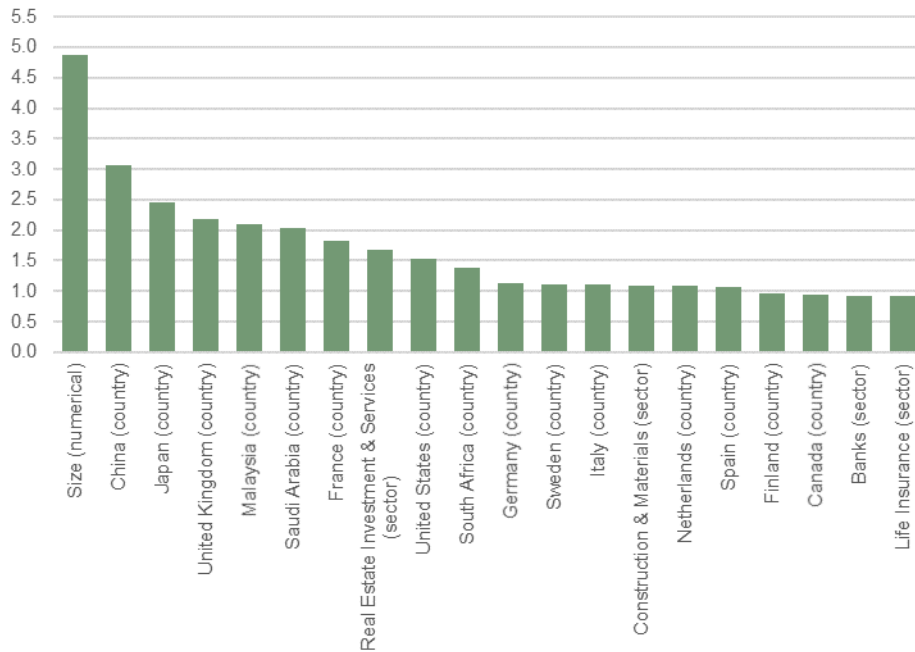
²¹ We perform the cross-validation procedure on 10 random sub-samples for each combination.

Appendix 2: a detailed study of the PLS regression outcomes

Variable Importance in Projection²² (VIP) scores summarize the contribution of each variable to the model and constitute the main measure of a variable's importance in a PLS framework.

Chart 24 plots VIP scores for the 20 first independent variables. As expected, firm size is by far the largest contributor to the model.

Chart 24. Independent variables from the PLS regression sorted by their importance in the model (measured by their VIP scores)



Sources: FTSE Russell, Beyond Ratings.

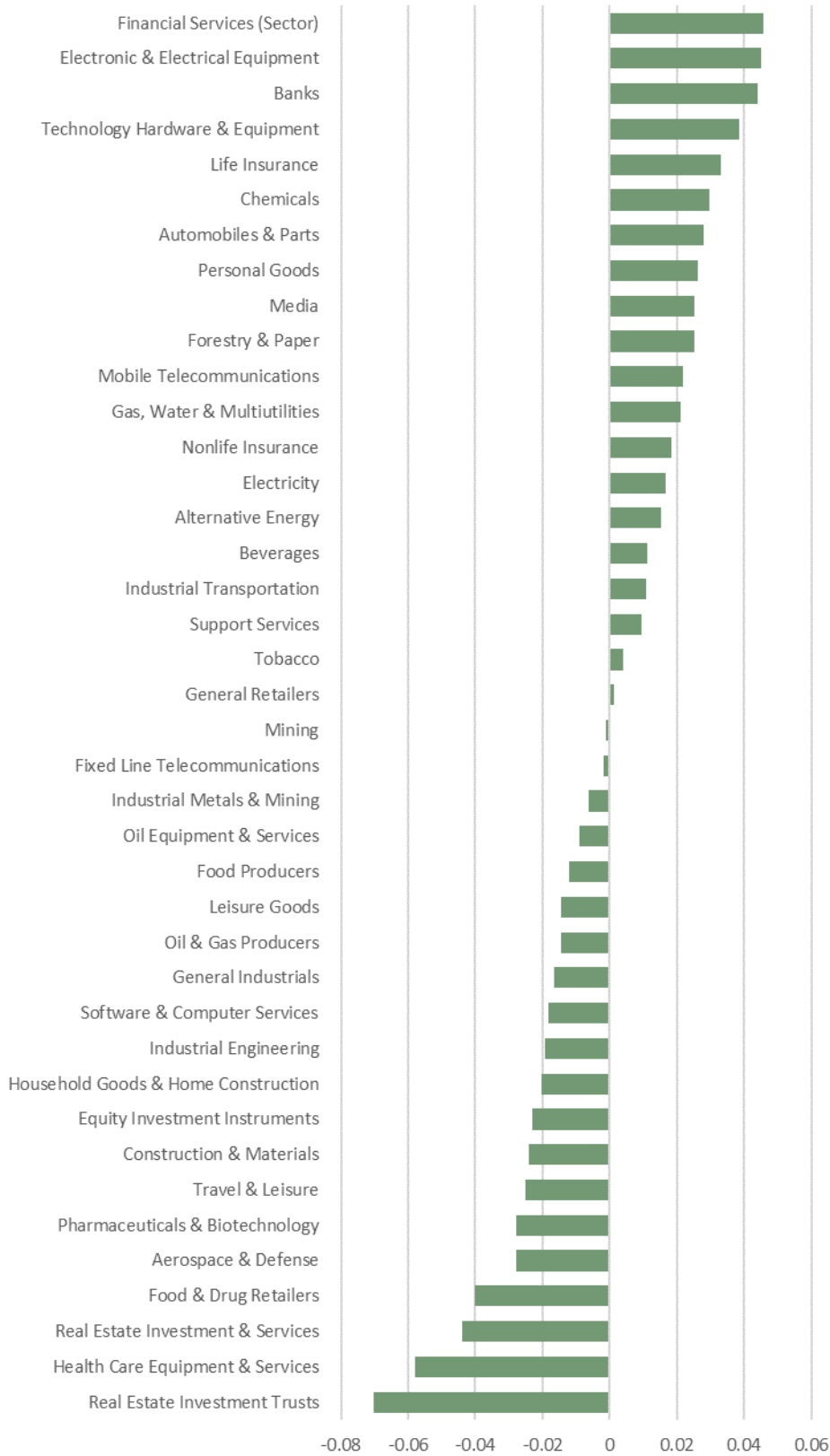
In the PLS framework, the higher the absolute value of the coefficient associated with the independent variable, the more this variable is related to the response variable. The sign displays the nature of the estimated relationship.

Thus, to study the link between each independent variable and the response one, we plot coefficients associated with each independent categorical variable.

Chart 25 shows coefficients for sector variables and Chart 26 for country ones.

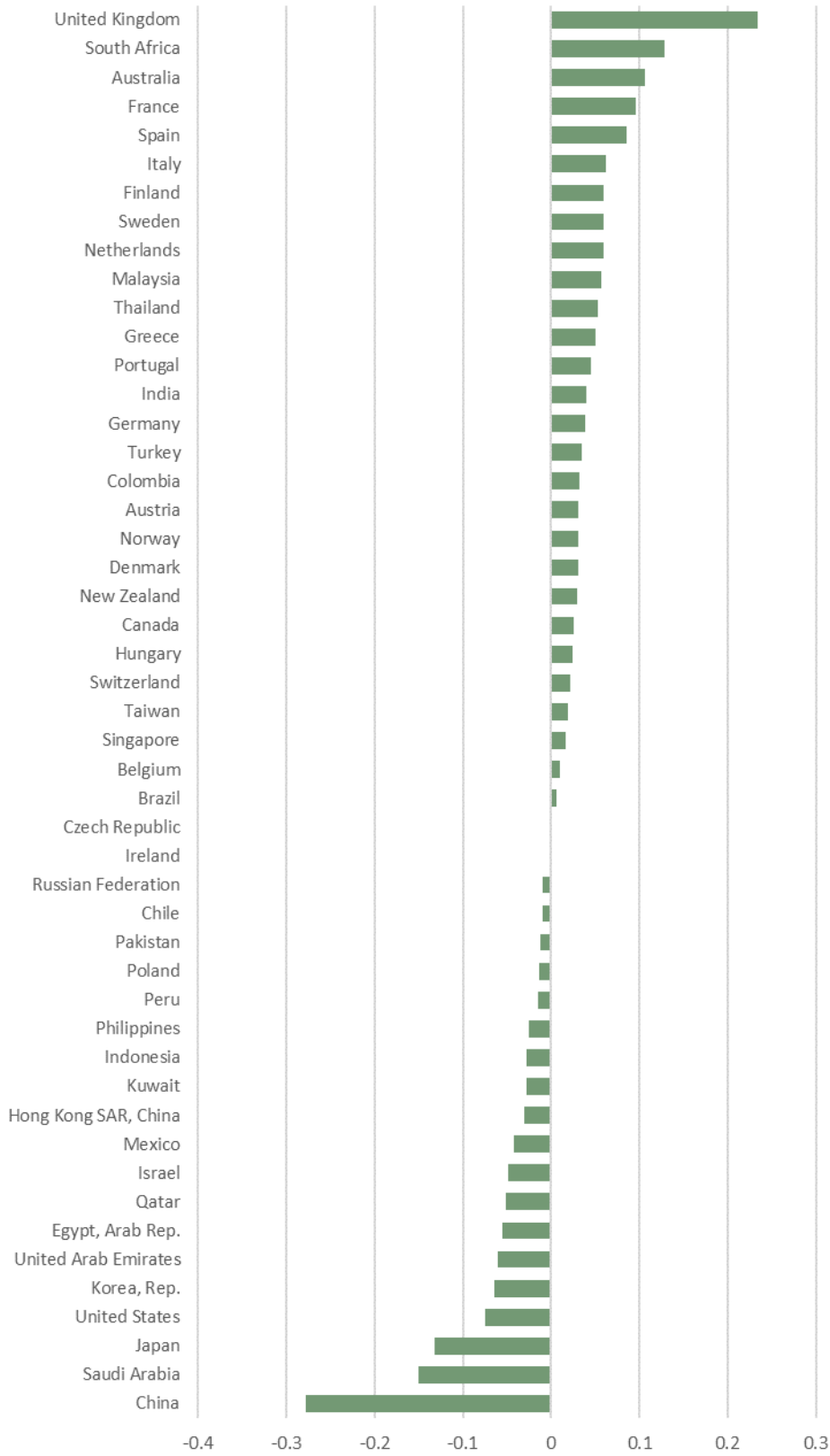
²² The VIP score of a variable is calculated as a weighted sum of the squared correlations between the PLS components and the original variable. The weights correspond to the percentage variation explained by the PLS component in the model. The number of terms in the sum depends on the number of PLS components found to be significant in distinguishing the classes. A variable with a VIP Score close to or greater than 1 (one) can be considered important in a given model.

Chart 25. Coefficients plot for sectors predictors in the PLS regression



Sources: FTSE Russell, Beyond Ratings.

Chart 26. Coefficients plot for countries predictors in the PLS regression



Sources: FTSE Russell, Beyond Ratings.

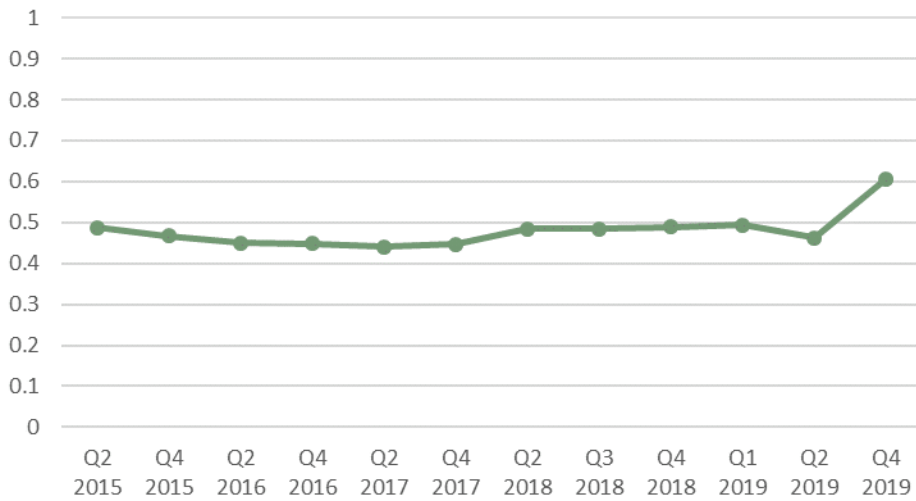
Appendix 3: outputs of the PLS regression's over time

The FTSE Russell ESG database provides ESG performance scores for about 4000 corporate issuers on average from September 2014 to December 2019 for the latest assessment.

To exploit the historical depth of the dataset and to check the robustness of our approach, we apply the optimized PLS regression algorithm to each historical set of companies' ESG scores.

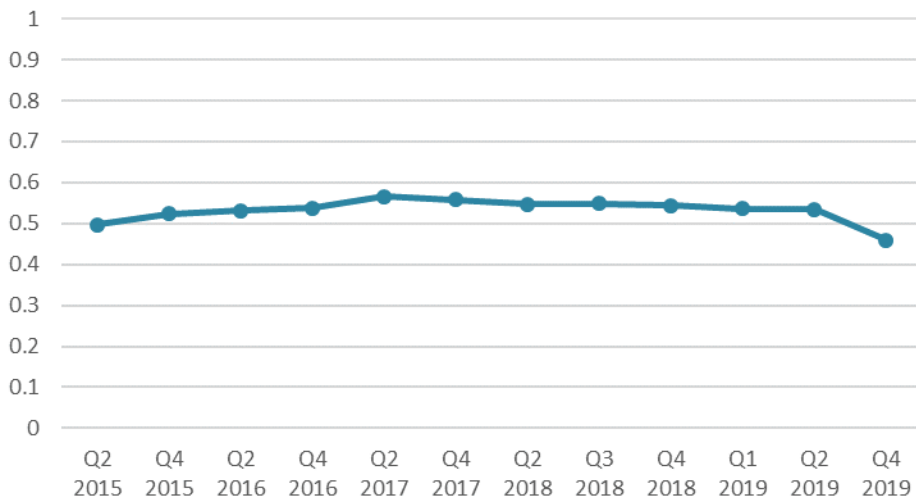
Chart 27 and Chart 28 show respectively the evolution of the r-squared and Mean Squared Error (MSE) metrics from the PLS regression. It shows a great stability over time but for an improvement of the model's quality on the latest assessment.

Chart 27. Evolution of the r-squared from the PLS regression over time



Sources: FTSE Russell, Beyond Ratings.

Chart 28. Evolution of the Mean Squared Error (MSE) from the PLS regression over time



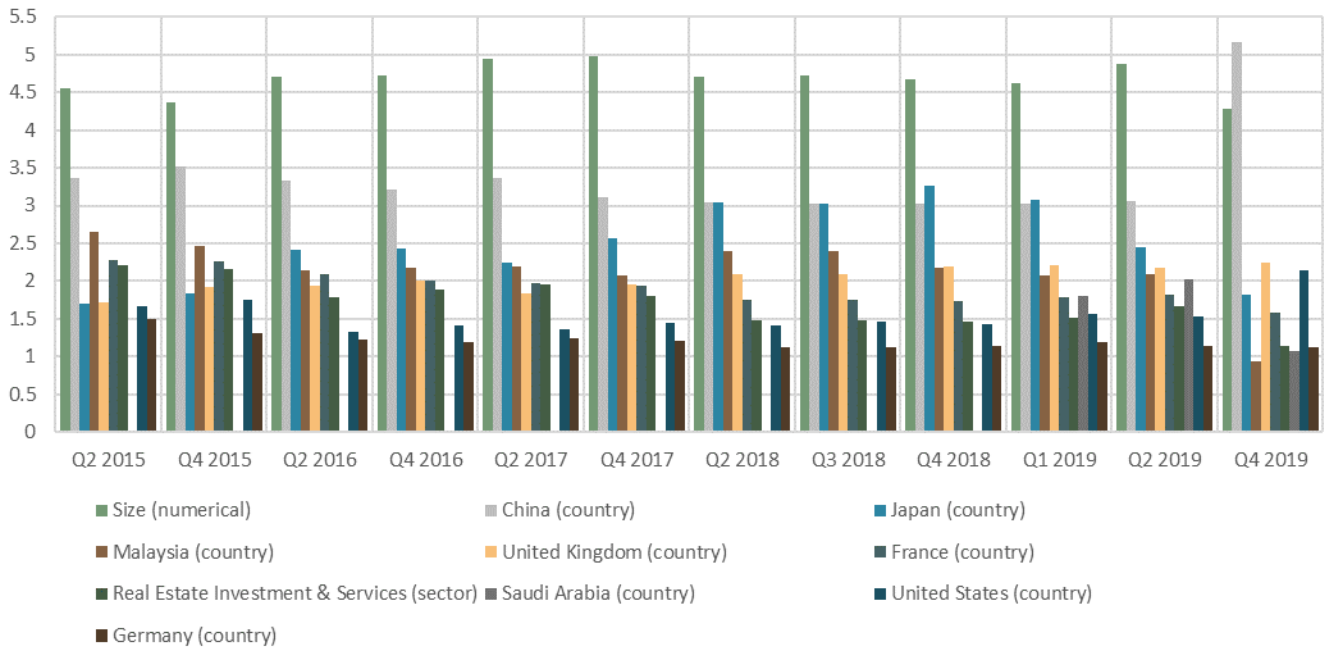
Sources: FTSE Russell, Beyond Ratings.

The improvement of the model's quality is due to the contribution of Chinese companies in the panel that have risen sharply on December 2019.

Indeed, as shown by Chart 28, while the VIP score from the PLS of the coefficient associated to China sets at 3 on average, it jumps to 5 on December 2019, to go beyond the size factor.

However, the order of the remaining independent variables remains nearly the same.

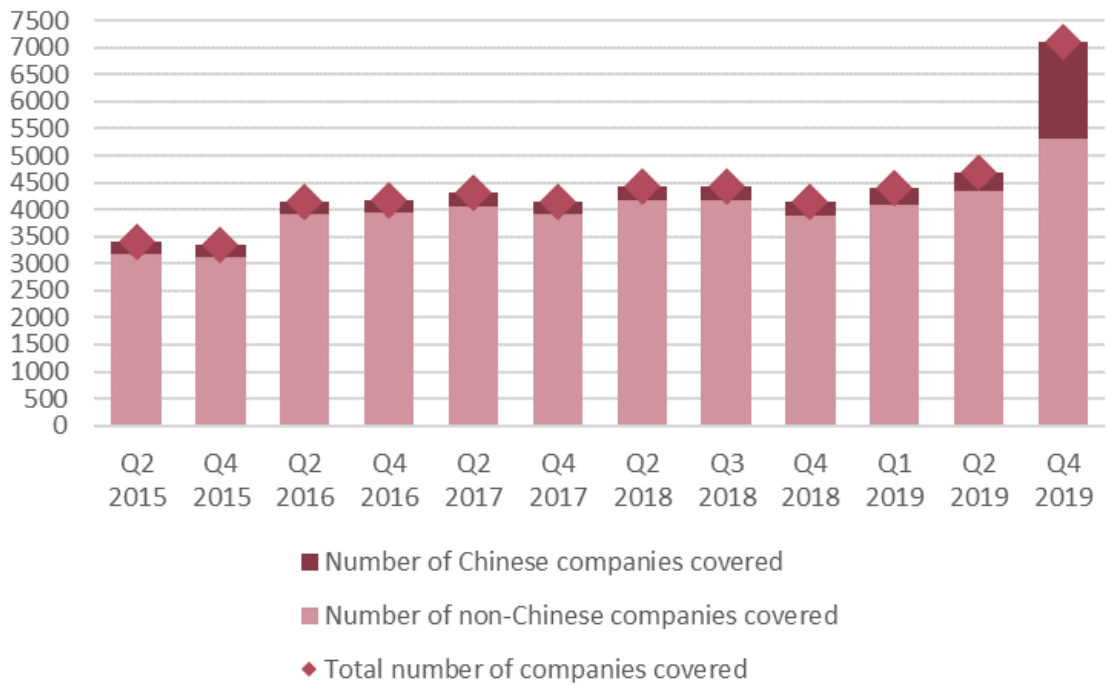
Chart 29. Evolution of the 10 first independent variables that contribute the most to the model (measured by their VIP scores)



Sources: FTSE Russell, Beyond Ratings.

This seems to be due to the increasing number of Chinese companies recently covered by the FTSE Russell ESG database, probably due to the inclusion of China A shares in most of FTSE Russell equity indices. While the database comprises +300 Chinese companies on average from 2017 to June 2019, they now account for more than 25% of the whole dataset (Chart 30).

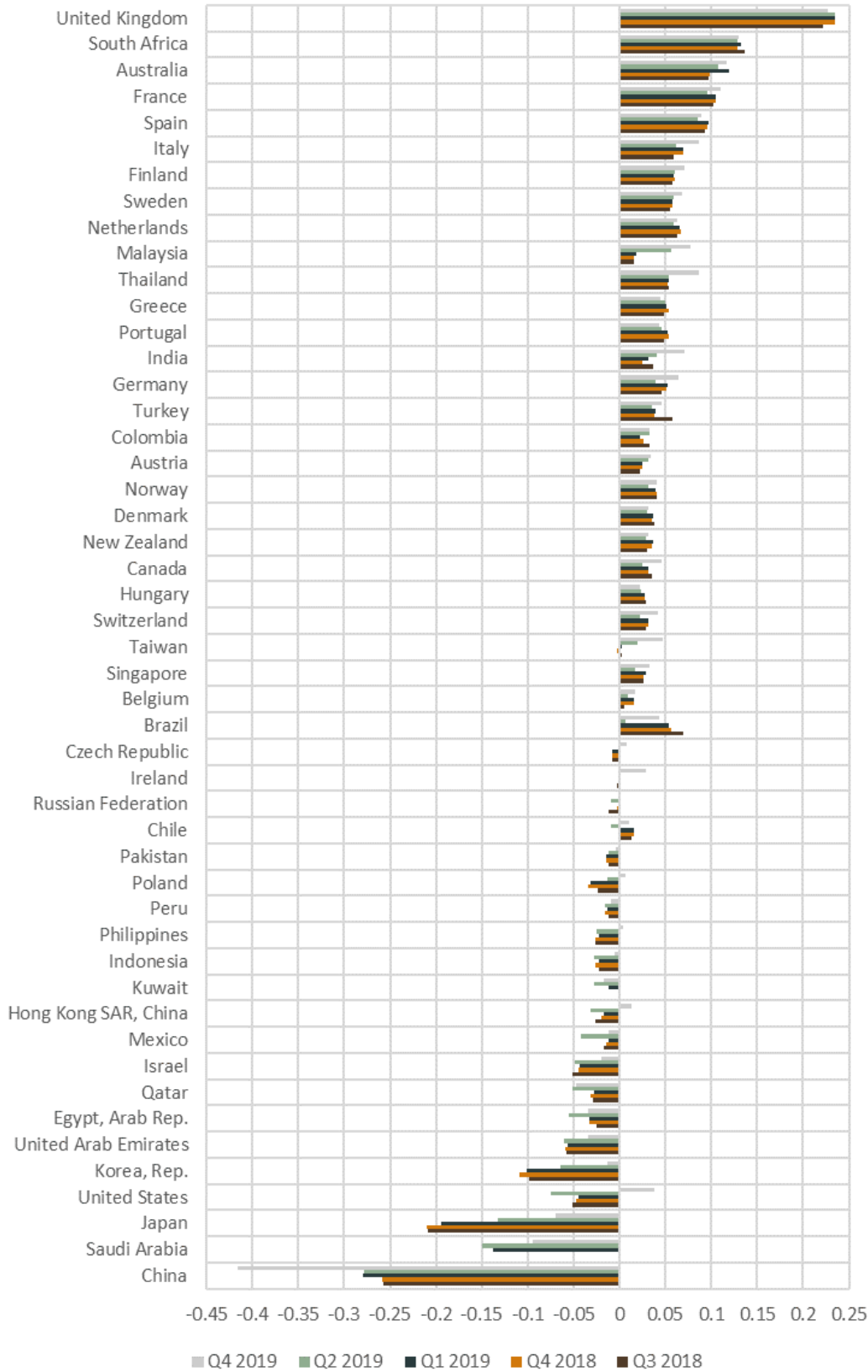
Chart 30. Number of Chinese companies covered by the FTSE Russell ESG database over time



Sources: FTSE Russell, Beyond Ratings.

However, as shown by Chart 31 below, this improvement does not have a significant impact on the level of the coefficients associated to most of the other countries (except for Hong Kong and Taiwan but the level of their coefficients is very low).

Chart 31. Evolution of the coefficients plot for countries predictors in the PLS regression

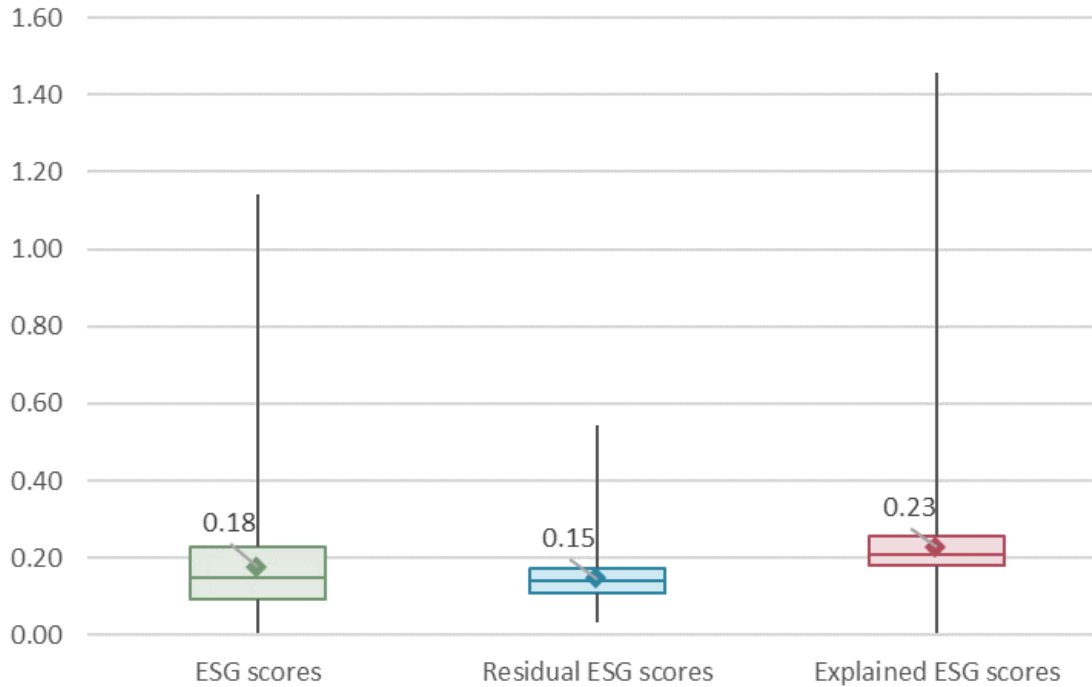


Sources: FTSE Russell, Beyond Ratings.

As a result, residual scores are stable over time, even more so than initial ESG scores.

Indeed, the average coefficient of variation²³ of residual scores' sets at 0.15 on average with a maximum of 0.54 while the initial scores have an average coefficient of variation of 0.18 and a maximum of 1.14 (Chart 32), which means that residual scores are stable over time, and even slightly more stable than the initial scores.

Chart 32. Main features of coefficient of variation* of ESG, Residual ESG and Explained ESG scores



Sources: FTSE Russell, Beyond Ratings. * We compute the coefficient of variation of a company if and only if it counts at least 9 historical data points.

²³ For a company, the coefficient of variation CV of its scores is defined as the ratio of the standard deviation σ to the average μ of its historical scores:
 $CV = \sigma/\mu$

About FTSE Russell

FTSE Russell is a leading global provider of benchmarks, analytics and data solutions with multi-asset capabilities, offering a precise view of the markets relevant to any investment process. For over 30 years, leading asset owners, asset managers, ETF providers and investment banks have chosen FTSE Russell indexes to benchmark their investment performance and create investment funds, ETFs, structured products and index-based derivatives. FTSE Russell indexes also provide clients with tools for performance benchmarking, asset allocation, investment strategy analysis and risk management.

To learn more, visit ftserussell.com; email info@ftserussell.com; or call your regional Client Service Team office

EMEA

+44 (0) 20 7866 1810

North America

+1 877 503 6437

Asia-Pacific

Hong Kong +852 2164 3333

Tokyo +81 3 4563 6346

Sydney +61 (0) 2 8823 3521

© 2020 London Stock Exchange Group plc and its applicable group undertakings (the "LSE Group"). The LSE Group includes (1) FTSE International Limited ("FTSE"), (2) Frank Russell Company ("Russell"), (3) FTSE Global Debt Capital Markets Inc. and FTSE Global Debt Capital Markets Limited (together, "FTSE Canada"), (4) MTSNext Limited ("MTSNext"), (5) Mergent, Inc. ("Mergent"), (6) FTSE Fixed Income LLC ("FTSE FI"), (7) The Yield Book Inc ("YB") and (8) Beyond Ratings S.A.S. ("BR"). All rights reserved.

FTSE Russell® is a trading name of FTSE, Russell, FTSE Canada, MTSNext, Mergent, FTSE FI, YB and BR. "FTSE®", "Russell®", "FTSE Russell®", "MTS®", "FTSE4Good®", "ICB®", "Mergent®", "The Yield Book®", "Beyond Ratings®" and all other trademarks and service marks used herein (whether registered or unregistered) are trademarks and/or service marks owned or licensed by the applicable member of the LSE Group or their respective licensors and are owned, or used under licence, by FTSE, Russell, MTSNext, FTSE Canada, Mergent, FTSE FI, YB or BR. FTSE International Limited is authorised and regulated by the Financial Conduct Authority as a benchmark administrator.

All information is provided for information purposes only. All information and data contained in this publication is obtained by the LSE Group, from sources believed by it to be accurate and reliable. Because of the possibility of human and mechanical error as well as other factors, however, such information and data is provided "as is" without warranty of any kind. No member of the LSE Group nor their respective directors, officers, employees, partners or licensors make any claim, prediction, warranty or representation whatsoever, expressly or impliedly, either as to the accuracy, timeliness, completeness, merchantability of any information or of results to be obtained from the use of FTSE Russell products, including but not limited to indexes, data and analytics, or the fitness or suitability of the FTSE Russell products for any particular purpose to which they might be put. Any representation of historical data accessible through FTSE Russell products is provided for information purposes only and is not a reliable indicator of future performance.

No responsibility or liability can be accepted by any member of the LSE Group nor their respective directors, officers, employees, partners or licensors for (a) any loss or damage in whole or in part caused by, resulting from, or relating to any error (negligent or otherwise) or other circumstance involved in procuring, collecting, compiling, interpreting, analysing, editing, transcribing, transmitting, communicating or delivering any such information or data or from use of this document or links to this document or (b) any direct, indirect, special, consequential or incidental damages whatsoever, even if any member of the LSE Group is advised in advance of the possibility of such damages, resulting from the use of, or inability to use, such information.

No member of the LSE Group nor their respective directors, officers, employees, partners or licensors provide investment advice and nothing contained herein or accessible through FTSE Russell products, including statistical data and industry reports, should be taken as constituting financial or investment advice or a financial promotion.

Past performance is no guarantee of future results. Charts and graphs are provided for illustrative purposes only. Index returns shown may not represent the results of the actual trading of investable assets. Certain returns shown may reflect back-tested performance. All performance presented prior to the index inception date is back-tested performance. Back-tested performance is not actual performance, but is hypothetical. The back-test calculations are based on the same methodology that was in effect when the index was officially launched. However, back-tested data may reflect the application of the index methodology with the benefit of hindsight, and the historic calculations of an index may change from month to month based on revisions to the underlying economic data used in the calculation of the index.

This document may contain forward-looking assessments. These are based upon a number of assumptions concerning future conditions that ultimately may prove to be inaccurate. Such forward-looking assessments are subject to risks and uncertainties and may be affected by various factors that may cause actual results to differ materially. No member of the LSE Group nor their licensors assume any duty to and do not undertake to update forward-looking assessments.

No part of this information may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the applicable member of the LSE Group. Use and distribution of the LSE Group data requires a licence from FTSE, Russell, FTSE Canada, MTSNext, Mergent, FTSE FI, YB, BR and/or their respective licensors.