

# Sustainable Investing: The Black Box of Environmental, Social, and Governance (ESG) Ratings

Subhash Abhayawansa and Shailesh Tyagi

## Subhash

### Abhayawansa

is an associate professor in accounting and finance at Swinburne University of Technology in Hawthorn, Australia.

[sabhayawansa@swin.edu.au](mailto:sabhayawansa@swin.edu.au)

## Shailesh Tyagi

is a climate change and sustainability specialist. He founded VIRIDI Global—a boutique consulting firm and worked as a strategic partner at GIST Advisory Switzerland.

[shailesh@viridiglobal.com](mailto:shailesh@viridiglobal.com)

## KEY FINDINGS

- There are significant divergences among the ratings and rankings provided by different ESG rating agencies.
- Differences among various ESG ratings and rankings are caused by differences in the definitions of ESG constructs (i.e., a theorization problem) and differences in the methods applied for measuring ESG performance of companies (i.e., a commensurability problem).
- Agencies providing ESG ratings and rankings are not transparent about what constitutes ESG performance and how ESG performance is measured, including information sources used.
- Theorization, commensurability, and transparency problems contribute to masking the true ESG risks and the performance of companies.

## ABSTRACT

Environmental, social, and governance (ESG) investing is becoming mainstream, and the COVID-19 pandemic has amplified the momentum. The interest in ESG investing creates greater demand for ESG data, ratings, and rankings, spawning a proliferation of agencies offering these products, which investors, academics, and regulators rely on unquestioningly. Research highlights that different ESG ratings and rankings produce significantly different assessments of the ESG performance of companies. This article examines the causes of differences in the ratings and rankings generated by different agencies. Findings indicate that the divergences among raters can be attributed to differences in the definitions of ESG constructs (i.e., a theorization problem) and methodological differences (i.e., a commensurability problem). While users of ESG ratings and rankings are advised to study the definitions and methodologies before their use, a lack of transparency about the data sources, weightings, and methodologies makes it difficult to ensure that companies' true ESG performance is accounted for when making portfolio selection and investment decisions. As a solution, the article notes that instead of attempting to compare and contrast ratings and rankings of different agencies, investors should determine the ESG constructs that are material to their own investment strategies and then match them with an ESG rating or ranking product that closely resembles those constructs.

## TOPICS

***ESG investing, information providers/credit ratings, portfolio construction, portfolio theory\****

\*All articles are now categorized by topics and subtopics. [View at PM-Research.com](#).

## OVERVIEW

The sustainable investing approach considers environmental, social, and governance (ESG) factors in portfolio construction and management. Once the domain only of socially responsible investors trying to exclude sustainability laggards from their portfolios, sustainability investing no longer is characterized as a niche approach. At the beginning of 2018, the assets under management using sustainable investing strategies<sup>1</sup> surpassed US\$30.7 trillion (or around 9% of global debt and equity) in the five major markets of Europe, the US, Canada, Japan, and Australia and New Zealand (Global Sustainable Investment Alliance 2018). In the 2 years leading up to 2020, professionally managed funds adopting sustainable investing strategies in the US alone increased by 42% to US\$17.1 trillion (US SIF 2020). The inflows to ESG investing funds continued to grow in 2020, achieving by the middle of the year the full-year amount for 2019<sup>2</sup> (Hale 2020). A 2019 survey of more than 550 CFA Institute members worldwide shows that more than half of the quantitative analysts or portfolio managers incorporate environmental and social factors into their investment analysis, and two-thirds factor in governance (Singh and Peters 2019).

The rise in sustainable investing has created a new industry that trades in ESG data and ratings on companies, funds dedicated to rated companies, and ESG index providers. Hawley (2017) notes more than 600 products from over 150 organizations that supply ESG data, ratings, and rankings; MSCI Inc. and Sustainalytics are the leading players. The industry is dynamic, with new ratings appearing and disappearing while rating organizations merge and realign at a rapid pace (Sadowski, Whitaker, and Buckingham 2010a). Academics, for instance, use ESG ratings to understand how the financial performance and market performance of sustainable companies differ from those of other firms—and what drives those differences. ESG ratings as well as academic research based on them can potentially influence regulators, impacting policies and regulations related to the disclosure and use of ESG data.<sup>3</sup> The unquestioning reliance on ESG ratings by various parties—including investors, academics, and regulators—makes it important to understand whether ESG ratings reflect the true ESG performance of companies and whether different ESG ratings converge. The purpose of this article is to identify the causes of the differences in the ratings and rankings produced by different agencies.

## ALL ESG RATINGS ARE NOT MADE EQUAL

Several studies and media articles point out that ESG ratings from different agencies display significant divergence (e.g., Chatterji et al. 2016; Yont, Jamie, and

---

<sup>1</sup>Sustainable assets under management include assets that are managed using one of the following seven strategies: (1) negative or exclusionary screening, (2) positive or best-in-class screening, (3) norm-based screening, (4) ESG integration, (5) sustainability-themed investing, (6) impact and community investing, and (7) corporate engagement and shareholder action (more details are provided in Global Sustainable Investment Alliance 2018).

<sup>2</sup>Flows are estimated for 315 sustainable open-end and exchange-traded funds available to US investors, which include equity, fixed-income, allocation, and alternatives funds that have an ESG, impact, or sustainable sector focus (Hale 2020).

<sup>3</sup>For example, the US Department of Labor (2020) new rule relating to the consideration of financial factors in selecting plan investments prohibits Employee Retirement Income Security Act (ERISA) plan fiduciaries from choosing investments to promote environmental, social, and other public policy goals unrelated to the interests of plan participants and beneficiaries. Also, recommendations of the SEC Investor Advisory Committee (2020) relating to ESG disclosure show that the growth in ESG data, ratings, and ranking providers was taken into consideration in the policy discussion about regulating ESG reporting.

Zhou 2018). Allen (2018) explains that in September 2018, FTSE placed Tesla last among the global automotive companies on ESG performance, but MSCI put it at the top, and Sustainalytics located it somewhere in the middle. When Semenova and Hassel (2015) explore the convergent validity of the environmental ratings produced by several core ESG rating agencies—MSCI, formerly known as Kinder, Lydenberg, and Domini Research & Analytics (KLD); Thomson Reuters' ASSET4 (now Refinitiv); and Global Engagement Services—they find that while the ratings have common dimensions, in aggregate, they do not converge. Dorfleitner, Halbritter, and Nguyen (2015) compare individual ESG and economic scores as well as the aggregate ESG scores of three rating products: Thomson Reuters ASSET4, MSCI/KLD ratings, and Bloomberg ESG data set. Using a subsample of companies covered by all three rating providers between 2002 and 2012, Dorfleitner, Halbritter, and Nguyen (2015) show that correlations between ASSET4 and Bloomberg are as high as 0.62 for the aggregate score and vary between 0.47 and 0.60 for individual dimensions whereas MSCI/KLD ratings display little resemblance to the other two ratings (with correlations varying between 0.05 and 0.39). Similarly, Chatterji et al. (2016) conclude that the six ratings they compare (i.e., MSCI/KLD, ASSET4, Calvert, FTSE4Good, DJSI, and Innovest) exhibit low convergence in their assessments of ESG factors. Furthermore, Hawley (2017) notes a correlation of merely 14% between the rankings based on the MSCI/KLD and Fortune Magazine's Best 100 Firms.

### Why Do Various ESG Ratings Differ?

Our review of the literature suggests that definitional and methodological differences contribute to the low convergence between different ESG ratings. Disagreements about the definitions of ESG factors, their composition, and the weighting for each of the factors (E, S, and G) are known as the theorization problem. Using a variety of approaches or methodologies to measure the ESG factors (even if definitions are agreed on) is termed the commensurability problem (Hawley 2017). The theorization problem originates in the different views held by rating agencies on the types of ESG factors that are considered to be financially material and on the degree to which they are viewed as material (i.e., the differences in their materiality maps). Dorfleitner, Halbritter, and Nguyen (2015, p. 256) identify the theorization problem in their comparison of the composition of ASSET4, MSCI/KLD, and Bloomberg ESG ratings; as an example of the theorization problem, they highlight that “animal testing is solely covered by ASSET4, whereas the question of whether a company's practice complies with environmental regulations is only considered by Bloomberg and KLD.” Chatterji et al. (2016) compare social ratings from six ESG raters, finding differences between ratings even when adjusted for the explicit variations in the definitions of ESG factors; thus, they demonstrate that the commensurability problem is independent of the theorization problem.

The commensurability problem relates to the differences in how the individual ESG dimensions are measured, especially the level of detail. One aspect of the commensurability problem relates to variations in the way that financially material ESG factors are measured, using indicators. Dorfleitner, Halbritter, and Nguyen (2015) explain that the three ratings they examine broadly cover the same aspects within the social dimension, and the indicators within it account for approximately half of the total data points and binary indicators. However, they note that the number of indicators dedicated to particular aspects within the social dimension, such as health and safety, vary significantly among the three ratings. The extent to which different raters conflate ESG impacts (which indicate ESG risks) and ESG performance contributes to low commensurability (Mattingly and Berman 2006)—and constitutes the other facet of the commensurability problem. Commenting on the environmental dimension

of ESG, the subject of their study, Mattingly and Berman (2006, p. 250) claim that “industry-specific risk drives EP [environmental performance] and that performance and risk are different constructs to be clearly separated.” Notwithstanding, Dorfleitner, Halbritter, and Nguyen (2015) find that ESG risk measures display the lowest level of correlation between different raters.

In addition, the commensurability problem also has origins in the differences in the information sources used in developing the ratings. The information sources vary, including company sustainability reports and other publicly disclosed information, surveys and interviews with the company, and independent information channels. Sadowski, Whitaker, and Buckingham (2010b) state that about one-half of the 120 raters in their examination rely only on public information sources while the rest employ either corporate self-disclosure alone or self-disclosure combined with public information sources. The use of information disclosed publicly by companies can bias the ratings in favor of both larger companies (because larger companies supply more disclosures than smaller companies after controlling for their level of corporate social performance) and companies domiciled in geographical regions such as Europe, which is characterized by more regulations and relatively mature traditions of sustainability reporting. Similarly, Sadowski, Whitaker, and Buckingham (2010b) identify a bias toward companies responding to information requests.

Moreover, the ratings of those agencies relying on publicly disclosed information, which is often based on annual reporting cycles, might not reflect the same period as the ratings produced by agencies that have access to more timely information sources. Furthermore, the questionable reliability and the lack of comparability of information disclosed in sustainability reports pose a problem which is created as a result of having neither commonly accepted standards for sustainability reporting nor a requirement to audit the reports. Also, depending on the sustainability reporting guidelines adopted, companies might disclose ESG information either with the investors in mind (e.g., if Sustainability Accounting Standards Board guidelines are implemented) or a multiplicity of stakeholders in mind (e.g., if Global Reporting Initiative Standards are applied). Hence, the same data source can provide different perspectives about different companies, complicating comparisons of ESG performance and risks across companies.

### Differences between ESG Ratings and Actual ESG Performance

Several limitations in the generally adopted rating methodologies contribute to a gap between the perceived performance (based on ESG ratings) and the true ESG performance and risks of companies. As the scores on various indicators are added to generate overall ESG ratings and ratings on each of the three factors (i.e., E, S, and G), trade-offs between indicators are inevitable (Hawley 2017). For instance, a company that scores exceptionally well on the diversity indicator might nonetheless perform poorly on board independence measures (for example, because either it has directors with longer tenures, a smaller board, or a smaller share of independent directors) (Yont, Jamie, and Zhou 2018). Most rating methodologies are likely to assign a similar weight to these two types of indicators and combine them—although a low level of board independence might constitute a greater risk than the absence of a gender-diverse board. Combining metrics masks important differences between companies in terms of performance and risk characteristics by aggregating empirically and conceptually distinct ESG constructs (Mattingly and Berman 2006).

Similarly, most raters fail to distinguish between disclosure (e.g., reported carbon emissions) and performance (e.g., actions to lower emissions) and thus disregard the need to weigh them differently (Hawley 2017). These problems are aggravated when one rater combines multiple components of a particular ESG factor into one indicator

(in contrast to one factor) while another rater separates them (Dorfleitner, Halbritter, and Nguyen 2015). Moreover, raters are criticized for not adequately focusing on material issues and trying to compare companies across industries that differ on material ESG issues (Sadowski, Whitaker, and Buckingham 2010a). Related to the determination of materiality is the time period over which given aspects of the three factors (E, S, and G) are considered to be material. For example, environmental elements such as loss of biodiversity or stranded assets resulting from climate change regulation might only be material in the long term whereas other transitional risks of climate change or releases of toxic waste might be material in the short term. Moreover, the time period over which a particular ESG factor is considered material can differ by industry. Rating agencies do not often disclose the time horizons employed in determining materiality, thus denying users information that is critical to understanding companies' true ESG risks and ESG performance. As a result of these limitations, ESG ratings are incapable of representing organizational realities. Allen (2018) argues, "Just because you can measure this stuff doesn't mean that you necessarily should."

### Exercise Caution

While the quality and comparability of ESG data and ratings remain hotly contested, investors are cautioned against overrelying on even top-level ESG scores (Chatterji et al. 2016; Yont, Jamie, and Zhou 2018). At the very least, the users of ESG data and ratings should understand what their chosen rater's methodology actually measures. Unfortunately, this advice, too, is problematic as most providers of ESG ratings and formulators of ESG rankings are not transparent—or are transparent only to a degree—about their data sources, weightings, and methodologies (Hawley 2017). For instance, some raters disclose categories and weightings, but not many explain how these categories and weightings are determined or the specific criteria applied within the categories (Sadowski, Whitaker, and Buckingham 2010a). The transparency problem is unlikely to recede because the rating methodologies (e.g., ESG indicators and weightings), data sources (e.g., interview participants), and measurement instruments (e.g., questionnaires) are proprietary and are viewed as the intellectual property of the raters. The usefulness of ESG ratings is further complicated by the potential conflicts of interest of some rating and ranking agencies because of their connections with the companies that they rate or that produce indexes based on their ratings.

### Concluding Remarks

The use of ESG ratings is tipped to increase, with the market starting to price company-specific ESG risks and to integrate ESG considerations alongside traditional financial analysis. Together with this growth, regulations are likely to follow on ESG disclosure, ESG ratings, and rankings and use of ESG factors in the investment processes (e.g., Securities and Exchange Commission 2020; US Department of Labor 2020). The theorization, commensurability, and transparency problems associated with ESG ratings are unlikely to disappear anytime soon. Thus, the only safeguard for investors lies in educating themselves on the types and quality of available ESG data, ratings, and rankings and on the strengths and limitations of the various methodologies and nuanced processes used by the raters (LaBella et al. 2019). We do not claim that ESG ratings are unusable; to the contrary, they can be valuable if applied with caution and as one of a number of inputs. The best way to embed ESG considerations in the investment process is by first determining what ESG constructs are material and then identifying an ESG rating that includes measures that closely resemble those constructs.

## REFERENCES

- Allen, K. 2018. "Lies, damned lies and ESG rating methodologies." *Financial Times*.
- Chatterji, A. K., R. Durand, D. I. Levine, and S. Touboul. 2016. "Do ratings of firms converge? Implications for managers, investors and strategy researchers." *Strategic Management Journal* 37 (8): 1597–1614.
- Dorflleitner, G., G. Halbritter, and M. Nguyen. 2015. "Measuring the level and risk of corporate responsibility—An empirical comparison of different ESG rating approaches." *Journal of Asset Management* 16 (7): 450–466.
- Global Sustainable Investment Alliance. 2018. *2018 Global Sustainable Investment Review*.
- Hale, J. 2020. "Sustainable Funds Continue to Rake in Assets During the Second Quarter." Accessed December 12, 2020. <https://www.morningstar.com/articles/994219/sustainable-funds-continue-to-rake-in-assets-during-the-second-quarter>.
- Hawley, J. 2017. *ESG Ratings and Rankings: All over the Map. What Does It Mean?* San Francisco, CA: Truevalue Labs.
- LaBella, M. J., L. Sullivan, J. Russell, and D. Novikov. 2019. "The Devil Is in the Details: The Divergence in ESG Data and Implications for Responsible Investing." New York: QS Investors.
- Mattingly, J. E., and S. L. Berman. 2006. "Measurement of Corporate Social Action: Discovering Taxonomy in the Kinder Lydenburg Domini Ratings Data." *Business & Society* 45 (1): 20–46.
- Sadowski, M., K. Whitaker, and F. Buckingham. 2010a. *Rate the Raters: Phase One—Look Back and Current State*. New York: SustainAbility.
- . 2010b. *Rate the Raters: Phase Two—Taking Inventory of the Ratings Universe*. New York: SustainAbility.
- SEC Investor Advisory Committee. 2020. *Recommendation of the SEC Investor Advisory Committee Relating to ESG Disclosure*. Washington, DC: U.S. Securities and Exchange Commission.
- Securities and Exchange Commission. 2020. *Modernization of Regulation S-K Items 101, 103, and 105*. Washington, DC: U.S. Securities and Exchange Commission.
- Semenova, N., and L. G. Hassel. 2015. "On the Validity of Environmental Performance Metrics." *Journal of Business Ethics* 132 (2): 249–258.
- Singh, M., and S. Peters. 2019. *The Case for Quarterly and Environmental, Social, and Governance Reporting*. New York: CFA Institute.
- U.S. Department of Labor. 2020. *Financial Factors in Selecting Plan Investments*. Washington, DC: U.S. Department of Labor.
- US SIF. 2020. *2020 Report on US Sustainable, Responsible and Impact Investing Trends*. Washington, DC: U.S. SIF, The Forum for Sustainable and Responsible Investment.
- Yont, C., A. Jamie, and M. Zhou. 2018. *Hard Decisions: Asia Faces Tough Choices in CG Reform*. Hong Kong: CLSA and Asian Corporate Governance Association.
- To order reprints of this article, please contact David Rowe at [d.rowe@pageantmedia.com](mailto:d.rowe@pageantmedia.com) or 646-891-2157.*