

ESG ratings: The mixed bag and its implications

Emma Sjöström^{a*}, Rickard Sandberg^{a,b} and Anders Westlund

^aMisum, Stockholm School of Economics, Stockholm, Sweden; ^bDepartment of Entrepreneurship, Innovation and Technology, Stockholm School of Economics, Stockholm, Sweden

*Correspondence details: emma.sjostrom@hhs.se, Stockholm School of Economics, Box 6501, SE-113 83 Stockholm

ESG ratings: The mixed bag and its implications

Asset owners and asset managers around the world are increasingly employing environmental, social and governance (ESG) criteria for investment analysis and decisions. Consequently, the market for providing corporate ESG ratings have proliferated. Research shows, however, that there is generally a low level of correlation between the ESG ratings of different ratings institutes. For example, the same company can receive a high rating from one institute and a medium or low by another. To gain a fuller understanding of the scope of the discrepancies and its implications for investors, corporations and academic researchers, this paper summarizes key academic literature on the divergence ESG ratings and offers a view on future research avenues.

Keywords: ESG ratings; ESG scores; divergence; discrepancy; corporate sustainability

Introduction

ESG ratings are designed to measure how well companies score on environmental, social, and governance (ESG) dimensions. This has become an essential input to many analysts and portfolio managers in order to identify ESG champions and laggards, whether it be for risk managing purposes or to fulfil an ESG investment mandate via for example a best-in-class approach (Wong and Petroy 2020). The ratings may also serve as an important input to the rated companies themselves, who might see a high rank as a testament to being ‘on the right track’ with its ESG efforts. Another key stakeholder group to the ratings industry is the many academic researchers who use it as input to their studies, for example to explore correlation between corporate social performance and corporate financial performance.

The ratings industry has grown in tandem with the steady increase of ESG oriented investors over the past two decades. Even large mainstream investors, such as Blackrock, State Street and Vanguard, now publicly commit themselves to integrating ESG into financial analyses and decision making, as do some of the world’s largest pension funds.

While ESG rankings have been instrumental for the inclusion of ESG into financial analysis and decision making (Wong and Petroy 2020), many studies have documented that

the same companies tend to be rated different by different institutes, and that there is generally low to moderate agreement between the different ESG ratings (Berg, Kolbel and Rigobon 2022; Christensen, Serafeim and Sikochi 2022; Billio et al. 2021; Bouyé and Menville 2021; Gibson, Krueger and Schmidt 2021; Lopez 2020; Widyawati 2020; Romero 2018; Chatterji et al. 2016; Dorfleitner, Halbritter and Nguyen 2015; Semenova and Hassel 2015). As each rating institute has developed its own proprietary measurement framework and rating method, it is perhaps not surprising that there will be some discrepancy in the results. At the same time, a certain level of uniformity might be necessary in order for the ratings to be trusted, and for investors and other key stakeholders to have a solid understanding of sustainability risk. It might also have implications for asset management, as the choice of ESG rating providers will render different investment universes and ultimately lead to different portfolio returns (Li and Polychronopoulos 2020).

This paper is the first to systematically summarize current literature on the topic of ESG ratings divergence, and has the aim to understand the following:

What drives the discrepancies, what are the main implications of the observed rating gaps, how might the ratings divergence be remedied, and what are the most pressing questions for future academic research in this area?

Discrepancies in ESG ratings

This paper is summarizing twelve scholarly manuscripts that empirically explore ESG ratings discrepancy. Eight of these were published in the period 2020-2022, indicating a quickly growing interest in this topic in the academic research community. The studies cover between three and seven rating agencies each (see Table 1 for an overview). Some of the most common ratings to appear in the studies include Sustainalytics, MSCI, Asset4 (now Refinitiv), and RobecoSAM (now S&P Global). The studied periods range from just one active trading day to a total of 12 years.

[Table 1 near here]

While it might be precarious to label specific correlation coefficients as ‘low’ or ‘moderate,’ we do purport that it would need be substantially higher than those observed in order to claim that there is only little discrepancy (as 1 would represent a perfect correlation). Aggregating data from six agencies, Berg, Kolbel and Rigobon (2022) record that the correlations are on average only 0.54 and varying from 0.38 to 0.71. Even lower average correlations are found in a study by Gibson, Krueger and Schmidt (2021) who reported an average of 0.45 between seven ratings, but with a maximum of 0.75. Lopez (2020), who are using machine learning, find across three rating agencies that pair-wise correlations are 0.65, 0.65 and 0.72 respectively. For more than 60 percent of the firms, however, the correlation is below 0.50, according to the same study. Zumente and Lace (2021), who have limited their study to European firms, find a lower highest mean correlation than both Lopez (2020) and Berg, Kolbel and Rigobon (2022) who both use a global sample. A correlation of 0.58 was recorded for the two most comparable ratings in the European sample. Widyawati (2020), focusing on internal consistency reliability, shows that during the period of 2009-2013 there was a low to moderate agreement for ESG ratings, reflecting differences in measurement concerns, and consistently an exceptionally low agreement of the ESG absolute scores.

Billio et al (2021) find low agreement between the four agencies covered, and a lack of common standards for how the E, S, and G are defined. There is a marked difference in the number of ESG indicators used by the agencies, ranging from 37 to 300. Additionally, the ratings companies are using different approaches in collecting as well as processing ESG data; as noted also by e.g. Dorfleitner, Halbritter and Nguyen (2015). Billio and colleagues also compare four ESG market indexes - STOXX, Dow Jones, Refinitiv and MSCI - and find that the agreement between them ranges from 35 to 59 percent. The overlap of the indexes’ common constituents is only 15 percent, corresponding to 48 constituents in total.

Dorfleitner, Halbritter and Nguyen (2015) find significant differences between the measurement concepts of ESG ratings from three main providers. The authors also include a risk dimension, by recording changes in ESG scores over time. They find that the different providers' ESG scores do not coincide to any large extent in terms of ESG risk, assessed as the percentage change of downward variations over time. To the list of studies with similar conclusions, we can add Chatterji and colleagues (2016) who also adjusts for differing definitions of CSR between the rating agencies (which does not resolve the discrepancy), Semenova and Hassel (2015) who focus specifically on the environmental dimension, and introduce a performance and risk focus, and Romero (2018), even if they are looking at only one active trading day - and seem to interpret this level of correlation in a more positive light than most of their academic peers.

Bouyé and Menville (2021) have a slightly different focus, looking at sovereign ESG ratings. They find a level of correlation of 0.72 to 0.95 in pair-wise comparisons. This indicates that sovereigns have much higher correlated ESG ratings than those on the corporate level. The authors suggest that this is expected, as sovereign data tends to come from more robust and standardized sources. Ratings divergence might however have several different causes, which we discuss in the next section.

What causes the divergence?

Berg, Kolbel and Rigobon (2022) identify three possible main sources of divergence between the institutes' ratings: The scope of categories, the measurement of categories, and the weights of categories. Comparing six different rating institutes, they find that it is particularly the difference in scope and measurement that drive the divergence (and less so the weights). The *scope* of categories refers to the different sets of attributes that are included. For example, one rating institute might include diversity, and another one not; some may include

lobbying, and others not. This can cause the ratings to diverge. The second source of divergence, *measurement* of categories, refers to different indicators for the same attribute. For example, to measure diversity, one rating institute might use gender diversity in the corporate board as an indicator, whereas another institute might instead use median age. The authors also note that indicators might differ in whether they focus on policies or outcomes. There also seems to be a 'raters effect' driving measurement discrepancy: A firm that receives high score in one category is more likely to also receive a high score in all other categories by the same rater. The *weight* divergence refers to when rating institutes are not aligned in the relative importance they assign to the attributes. There is a substantial difference between raters here, and the top three most important categories for the raters are rarely the same. Yet, the weights do not account for divergence to the same extent as scope and measurement.

Several authors have also sought to detangle the E, S and G dimensions, with varying results. Some authors find that that governance factors are driving the divergence to a larger extent than environmental and social dimensions (Berg, Kolbel and Rigobon 2022; Gibson, Krueger and Schmidt 2021; Widyawati 2020). For example, one study records that corporate governance ratings have a correlation of on average 0.33, while social dimensions average 0.42 and environment 0.53 (Berg, Kolbel and Rigobon 2022). Similar results are found by Gibson, Krueger and Schmidt (2021) with the correlation for governance at 0.19, and for environment 0.43. Dorfleitner, Halbritter and Nguyen (2015) find the lowest pairwise rater correlations for the environmental dimension and governance, the lowest being 0.04 and 0.05 respectively (likely caused by the particularities of KLD). Christensen, Serafeim and Sikochi (2022), who are particularly interested in the role of ESG disclosure for ESG ratings divergence, find that it is rather the environmental and social dimensions that drive the divergence. The authors say that this is perhaps not surprising, given that the discussion on

what constitutes good corporate governance has been ongoing for far longer than the attempts to streamline environmental and social performance indicators. The fact that authors find different results is a natural outcome of the fact that they compare different sets of rating agencies, and we can (apparently) not draw a general conclusion about the environmental, social or governance factor is generally more or less correlated across rating agencies.

Some studies also take a sector perspective. In Lopez's (2020) sample, the average agreement is lowest in the Energy sector (0.55) and highest in Technology (0.77), Cyclical Consumer Goods & Services (0.74), and Financials (.74). The authors suggest that the sectors with higher agreement place less emphasis on environmental factors, which could help explain the results. Zumante and Lace (2021) find the most pronounced lack of convergence for Automobiles and components, Media and entertainment, Technology, and Utilities, but hesitate to draw conclusions from these results. Gibson, Krueger and Schmidt (2021) record the highest ratings disagreement for environmental and social dimensions in the Consumer durables and Telecommunications industries, and for the governance dimension in Financial services companies.

Overall, there seems to be stronger agreement for the lowest ranked companies: Lopez and colleagues (2020) report a correlation of 0.95 for the 10 percent of firms that are the worst-performing on ESG. Furthermore, studies find that the disagreement is higher for larger firms, and for firms that do not have a credit rating, firms with better ESG disclosure, and firms with greater analyst following; while more profitable firms tend to have lower disagreement in the ESG ratings (Christensen, Serafeim and Sikochi 2022; Gibson, Krueger and Schmidt 2021).

Some studies have also examined if the level of disclosure has any impact on how much ratings differ. It seems that generally, ESG disclosure is negatively correlated with ESG ratings discrepancy (Christensen, Serafeim and Sikochi 2022; Romero 2018). In other

words, the better the disclosure, the less do rating institutes agree on the ESG score. This may be explained by the fact that greater disclosure demands a higher level of subjectivity, in order to assess whether the disclosure means higher or lower performance. This subjectivity causes divergence; the more disclosure there is, the more information to disagree about. Christensen, Serafeim and Sikochi (2022) find a U-shaped relationship: The highest disagreement is with the firms that exhibit the lowest ESG ratings, followed by those who have the highest ESG ratings. Firms in the middle of the ranking distribution are those that have the lowest levels of disagreement. Christensen, Serafeim and Sikochi (2022) comment that this suggests that rating institutes disagree more when the stakes are high, as most investors are interested in either the top-performers for a positive screening approach, or the worst-performers for negative screening and engagement approaches. Furthermore, Christensen, Serafeim and Sikochi (2022) find that ESG ratings disagreement tends to expand, relative to control firms, after a country or a stock exchange implements mandatory ESG disclosure requirements. Looking at disclosure from another angle, namely whether corporate disclosure levels correlate with ESG scores, Romero (2018) find that the higher the ESG disclosure of companies the higher ESG score they tend to receive. In this context, it is notable that research has found a positive correlation between firm size, a company's available resources for providing ESG data, and the availability of a company's ESG data on the company's sustainability performance (Drempetic, Klein and Zwergel 2020). This implies that current ESG score measures favor larger firm with more resources.

Billio et al (2021) explore whether the geographical allocation of the index constituents may explain a higher or lower agreement rate but finds only a marginal impact; Berg, Kolbel and Rigobon (2022) are also ruling out constituents' geographical location as a driver for divergence. Geography may however come in to play in terms of the raters' (rather than the constituents') location: Where raters are situated in the world may influence their

theorization of CSR, which in turn will influence the ratings. Chatterji et al (2016) notes that KLD, based in the US, puts more weight on social issues than the European agency Asset 4:

The average tetrachoric correlations between U.S. raters (0.45) and between EU raters (0.53) are higher than the average correlation between all raters (0.31), providing suggestive evidence that geographically proximate raters may have closer theorizations and/or higher commensurability of CSR. (Chatterji et al 2016, 1604).

This line of argument is also forwarded by Eccles and Strohle (2018), who find that the rating agencies' social and contextual origins strongly influences their conception of sustainability and ultimately the measurement of ESG.

Constructing ESG ratings is in other words not only a technical matter; it is also a contextual one. How ESG is interpreted and translated may depend on the rating institutes motivations, cultural heritage, national context, and other such factors (Eccles and Stroehle 2018). These circumstances may be a potent explainer for ratings divergence. A major divider is whether the ratings institutes are primarily driven by a value-based or values-based motive (Eccles and Stroehle 2018). Value-based ratings would seek to capture how ESG impacts financial performance, whereas values-based ratings are primarily seeking to score how companies impact ESG. It follows from this that categories, indicators and weights will be different, depending on the underlying motive, or even the institutional logics that guide their behaviour (Sandberg and Sjöström 2020). We must therefore understand ESG ratings as not only a technical matter but also as social constructs.

Implications of ratings divergence

Academic literature conveys a united picture of an ESG rating industry that produces divergent results about corporate ESG performance. Such discrepancies implicate that investors will be faced with uncertainty about companies' actual performance and thus will have a less robust ground for risk evaluation.

Scope divergence may be unavoidable, as the different facets of sustainability and ESG can be interpreted in various ways (Berg, Kolbel and Rigobon 2022; Eccles and Stroehle 2018). However, Berg, Kolbel and Rigobon (2022) purport that the discrepancy they have found across the rating institutes will make it difficult for companies to know what is expected by them and what the market appreciates, when the same effort is praised by some rating institutes and not by others. Additionally, research has found that that rating disagreement hinders the incorporation of value-relevant ESG news into prices (Serafeim and Yoon 2022). This ambiguity may hamper corporate ESG efforts. It might also mean that the trillions of dollars that are being invested based on ESG criteria, where ESG ratings is one key input, are being channeled to or from companies on grounds that are not entirely clear to the investors themselves.

Another effect of the rating discrepancy might be, according to the same authors, that ESG efforts are not reflected in corporate stock and bond prices. The influence of investor taste on asset prices will only materialize if a large-enough fraction of investors has a uniform preference.

For the academic community that is leaning on ESG ratings as a proxy for company ESG performance, research results might differ depending on which rating is used, and results might not be fit for generalization. For example, ESG ratings have been extensively used in studies as a proxy for the success of shareholder resolutions and the relation between corporate financial performance and corporate social performance. We also note that KLD seems to be a widely used rating for academic studies (Eccles, Lee and Stroehle 2020) while at the same time the discrepancy between KLD and other ratings is the largest (e.g. Berg, Kolbel and Rigobon 2022; Dorfleitner, Halbritter and Nguyen 2015). To overcome this, it is suggested that academic researchers include more than one ESG rating in their data analysis, or to use just one ESG rating for the purposes of measuring a specific company property

(such as social capital) as long as researchers can motivate the choice of measurement approach and aggregation procedure for their study (Berg, Kolbel and Rigobon 2022).

As a remedy to the ESG ratings divergence, many researchers call for standardization of ESG data (Billio et al. 2021; Lopez 2020; Romero 2018; Escrig-Olmedo et al. 2019). The lack of a standardized definition of ESG – and CSR before that – is an issue that is relevant not only in the ratings setting but also more generally. Competing standards for ESG reporting, ESG labelling of investment products, et cetera, might however be sub optimal for investors. It increases the level of subjectivity and undermines trust.

Others point out that, on the methods side, harmonization is not necessarily an end goal (Bouyé and Menville 2021; Lopez 2020). Different modelling choices will generate different results, and in a markets-based environment different approaches should be embraced (Eccles and Strohle 2018). On one condition: That the different choices made by the agencies are transparent. If it was easy for investors, academic researchers, and other stakeholders to know the details of the methodologies and aggregation procedures, and the premises upon which they were designed, ratings divergence might not be a major concern. Clearly, however, the ratings industry is opaque (Berg, Kolbel and Rigobon 2022; Billio et al. 2021; Lopez 2020; Widyawati 2020; Dorfleitner, Halbritter and Nguyen 2015). Researchers also note that the ratings methodologies and assessment criteria have a tendency to change with time, making it even more difficult for external stakeholders to evaluate and understand the underlying work (Berg, Fabisik and Sautner 2021; Escrig-Olmedo et al. 2019).

Given the relative novelty of ESG ratings, it is however not surprising that there is an array of approaches to constructing appropriate methodologies. ESG ratings can be contrasted with credit ratings, which have a much higher convergence. This convergence did not happen overnight (Bouyé and Menville 2021), and the same can be said for the general formalization of financial analysis (Christensen, Serafeim and Sikochi 2022). For credit ratings, the

convergence was supported by regulation and by the development of standardized data (Bouyé and Menville 2021). The credit ratings market has with time become highly concentrated, with only three major players today. We currently witness similar developments for ESG ratings agencies (Wong and Petroy 2020), with a lot of mergers and acquisition activity in the past few years (rightfully predicted by Escrig-Olmedo et al. 2010; see also Avetisyan and Hockerts (2017) for more on the driving forces and impact of these consolidations). This could potentially also contribute towards mainstreaming ESG scores. Indeed, several ESG ratings institutes compare themselves, in their marketing material, to traditional financial institutes (Chatterji et al. 2016), likely as a way to increase their legitimacy.

From an investor perspective, ratings can indeed cause frustration. A study by the consultancy SustainAbility reports:

Investors interviewed expressed strong critiques of ratings, from inaccuracies and use of old or backwards-looking data, to more fundamental concerns about whether ESG performance can ever be distilled into a single score. Investors also complained about holes in corporate data and the need for companies to improve disclosure, reporting and transparency. Almost all the investors we engaged believe that their own research analysts know better than ESG rating analysts, often because ratings research teams don't have as much experience and are spread too thin. (Wong and Petroy 2020,14)

Avenues for future research

In this paper, we have summarized recent literature that have compared ESG ratings and commented on the implications of the rating discrepancy that has now been widely documented. These studies present a number of suggestions for future research in the field of ESG ratings discrepancy, such as separating the different constructs of performance and risk (Semenova and Hassel 2015), exploring what drives the 'rater effect,' that is the tendency that a firm that receives high score in one category is more likely to also receive a high score

in all other categories by the same rater, and whether incentive structures play a role (Berg, Kolbel and Rigobon 2022), and the effects of different ratings on the relationship between corporate social performance and financial performance in order to resolve some of that debate (Widyawati 2020).

As the divergence is already well documented, we would also argue that more research efforts need to be directed towards *refining the methods* for company ratings. In that vein, we suggest that a more stakeholder-oriented approach is warranted, which better accounts for the user perspective. We would also propose to explore a model where the input to the ESG scoring is based on the surveyed perceptions of those who regularly evaluate companies, namely the financial analysts and portfolio managers. This might also remedy the investor concern that we cited in the previous section, that ‘(a)lmost all the investors we engaged believe that their own research analysts know better than ESG rating analysts.’ Additionally, it appears that much of the current research and commercial products on ESG scores and rankings are merely descriptive or are not using scientifically proven measurement methods. Academic research therefore has a key role in advancing and complementing the models and frameworks that are available today.

We would also welcome a sociological take on ESG-ratings, as discrepancies, but even more so the dissatisfaction with the discrepancies, raises questions about the quest for accuracy, the notion of ‘truth in numbers,’ and the urge to represent and organize complex social realities through numbers (Miller 2001; Pérezts, Andersson and Lindebaum 2021). Such research could further stimulate a discussion about the continuous process among financial market actors to frame ESG in such a way that it could ‘fit’ into the existing format but also logic of financial analysis and communication, i.e. as part of the financialization of ESG (Beunza and Ferraro 2019; Giamporcaro and Gond 2016; Arjaliès 2010). It could also feed into a vein of more critical research, where ESG might be seen as being ‘colonized’ by

financialized valuations (Chiapello 2015), as well as a concern about pricing the priceless, such as the natural environment (Fourcade 2011) or other dimensions of ESG.

Today's ESG ratings are not encouraging neither investors nor companies to reflect on or address 'negative externalities', or to take a more system-wide approach to ESG and corporate sustainability (Crona, Folke and Galaz, 2021). This would arguably be necessary if we are to reach global sustainability goals, which is for many investors and companies an important motivation for including an ESG perspective in the first place (alongside financial value-based motives). Recent research has proposed that risk frameworks need to take into account 'aggravation risk' which is 'externalities created by one industry, which contribute to large-scale environmental change that comes back and affects the sector itself (...)' (Crona, Folke and Galaz 2021, 620). Such focus on externalities and earth-system perspectives related to financial analysis and corporate activity would also merit attention in future research about ESG ratings.

Acknowledgements

Funding details

This work was supported by the Nasdaq Nordic Foundation, and Mistra - the Swedish foundation for strategic environmental research.

Disclosure statement

The authors report there are no competing interests to declare.

References

- Arjaliès, D.-L. 2010. "A Social Movement Perspective on Finance: How Socially Responsible Investment Mattered." *Journal of Business Ethics* 92 (S1): 57–78. doi:[10.1007/s10551-010-0634-7](https://doi.org/10.1007/s10551-010-0634-7).
- Avetisyan, E., and K. Hockerts. 2017. "The Consolidation of the ESG Rating Industry as an Enactment of Institutional Retrogression." *Business Strategy & the Environment* 26 (3): 316–330. doi:[10.1002/bse.1919](https://doi.org/10.1002/bse.1919).
- Berg, F., K. Fabisik, and Z. Sautner. 2021. "Rewriting History II: The Unpredictable Past of ESG Ratings" *ECGI, Finance Working Paper No. 708/2020*.
- Berg, F., J. F. Kölbel, and R. Rigobon. 2022. "Aggregate Confusion: The Divergence of ESG Ratings." *Review of Finance* 26 (6): 1315–1344. doi:[10.1093/rof/rfac033](https://doi.org/10.1093/rof/rfac033).
- Beunza, D., and F. Ferraro. 2019. "Performative Work: Bridging Performativity and Institutional Theory in the Responsible Investment Field." *Organization Studies* 40 (4): 515–543. doi:[10.1177/0170840617747917](https://doi.org/10.1177/0170840617747917).
- Billio, M., M. Costola, I. Hristova, C. Latino, and L. Pelizzon. 2021. "Inside the ESG Ratings: (Dis)Agreement and Performance." *Corporate Social Responsibility & Environmental Management* 28 (5): 1426–1445. doi:[10.1002/csr.2177](https://doi.org/10.1002/csr.2177).
- Bouyé, E., and D. Menville. 2021. "The Convergence of Sovereign ESG Ratings". Policy Research Working Paper No. 9583. <http://hdl.handle.net/10986/35291>. Washington, DC: World Bank.
- 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. doi:[10.1002/smj.2407](https://doi.org/10.1002/smj.2407).
- Chiapello, E. 2015. "Financialisation of Valuation." *Human Studies* 38 (1): 13–35. doi:[10.1007/s10746-014-9337-x](https://doi.org/10.1007/s10746-014-9337-x).
- Christensen, D. M., G. Serafeim, and A. Sikochi. 2022. "Why Is Corporate Virtue in the Eye of the Beholder? The Case of ESG Ratings." *Accounting Review* 97 (1): 147–175. doi:[10.2308/TAR-2019-0506](https://doi.org/10.2308/TAR-2019-0506).
- Crona, B., C. Folke, and V. Galaz. 2021. "The Anthropocene Reality of Financial Risk." *One Earth* 4 (5): 618–628. doi:[10.1016/j.oneear.2021.04.016](https://doi.org/10.1016/j.oneear.2021.04.016).
- Dorfleitner, G. H., 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.
doi:[10.1057/jam.2015.31](https://doi.org/10.1057/jam.2015.31).
- Drempetic, S., C. Klein, and B. Zwergel. 2020. “The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review”. *Journal of Business Ethics* 167: 1–18. doi:[10.1007/s10551-019-04164-1](https://doi.org/10.1007/s10551-019-04164-1)
- Eccles, R. G., and J. C. Strohle. 2018. *Exploring Social Origins in the Construction of ESG Measure*[Working paper]. Saïd Business School, University of Oxford.
- Eccles, R. G., L.-E. Lee, and J. C. Strohle. 2020. “The Social Origins of ESG: An Analysis of Innovest and KLD.” *Organization & Environment* 33 (4): 575–596.
doi:[10.1177/1086026619888994](https://doi.org/10.1177/1086026619888994).
- Escrig-Olmedo, E., M.A. Fernández-Izquierdo, I. Ferrero-Ferrero, J.M. Rivera-Lirio, and M.J. Muñoz-Torres. 2019. “Rating the Raters: Evaluating how ESG Rating Agencies Integrate Sustainability Principles.” *Sustainability* 11 (915): 1-16.
doi:[10.3390/su11030915](https://doi.org/10.3390/su11030915)
- Escrig-Olmedo, E., M.J. Muñoz-Torres, and M.A. Fernandez-Izquierdo. 2010. “Socially Responsible Investing: Sustainability Indices, ESG Rating and Information Provider Agencies.” *International journal of sustainable economy*, 2 (4): 442-461.
doi:[10.1504/IJSE.2010.03549](https://doi.org/10.1504/IJSE.2010.03549)
- Fourcade, M. 2011. “Cents and Sensibility: Economic Valuation and the Nature of Nature.” *American Journal of Sociology* 116 (6): 1721–1777. doi:[10.1086/659640](https://doi.org/10.1086/659640).
- Giamporcaro, S., and J.-P. Gond. 2016. “Calculability as Politics in the Construction of Markets: The Case of Socially Responsible Investment in France.” *Organization Studies* 37 (4): 465–495. doi:[10.1177/0170840615604498](https://doi.org/10.1177/0170840615604498).
- Gibson Brandon, R. K., P. Krueger, and P. S. Schmidt. 2021. “ESG Rating Disagreement and Stock Returns.” *Financial Analysts Journal* 77 (4): 104–127.
doi:[10.1080/0015198X.2021.1963186](https://doi.org/10.1080/0015198X.2021.1963186).
- Global Sustainable Investment Alliance. 2018. *Global Sustainable Investment Review*.
http://www.gsi-alliance.org/wp-content/uploads/2019/06/GSIR_Review2018F.pdf
vol. 2018.
- Li, F., and A. Polychronopoulos. 2020. “What a Difference an ESG Ratings Provider Makes.” *Research Affiliates* 15.
- Lopez, C. C. 2020. “ESG Ratings: The Road Ahead”.
https://milkeninstitute.org/sites/default/files/reports-pdf/ESG_Ratings_0.pdf. Milken Institute.

- Miller, P. 2001. "Governing by Numbers: Why Calculative Practices Matter." *Social Research*: 379–396.
- Pérezts, M., L. Andersson, and D. Lindebaum. 2021. "Numbers and Organization Studies: Book Review Symposium Editorial." *Organization Studies* 42 (8): 1351–1356. doi:[10.1177/0170840621989214](https://doi.org/10.1177/0170840621989214).
- Romero, S. J. 2018. "Using ESG Ratings to Build a Sustainability Investing Strategy." *CPA Journal*.
- Sandberg, J. and E. Sjöström. 2020. *Motivations for Investment in Sustainable Consumption and Production*. In: Bali Swain, R. and S. Sweet. (eds.) *Sustainable Consumption and Production, Volume I*. Palgrave MacMillan.
- Semenova, N., and L. G. Hassel. 2015. "On the Validity of Environmental Performance Metrics." *Journal of Business Ethics* 132 (2): 249–258. doi:[10.1007/s10551-014-2323-4](https://doi.org/10.1007/s10551-014-2323-4).
- Serafeim, G., and A. Yoon. 2022. "Stock Price Reactions to ESG News: The Role of ESG Ratings and Disagreement." *Review of Accounting Studies*. doi:[10.1007/s11142-022-09675-3](https://doi.org/10.1007/s11142-022-09675-3).
- Widyawati, L. 2021. "Measurement Concerns and Agreement of Environmental Social Governance Ratings." *Accounting & Finance* 61 (S1): 1589–1623. doi:[10.1111/acfi.12638](https://doi.org/10.1111/acfi.12638).
- Wong, C., and E. Petroy. 2020. "Rate the Raters 2020: Investor Survey and Interview Results." <https://www.sustainability.com/globalassets/sustainability.com/thinking/pdfs/sustainability-ratetheraters2020-report.pdf> SustainAbility.
- Zumente, I., and N. Lāce. 2021. "ESG Rating—Necessity for the Investor or the Company?." *Sustainability* 13 (16): 8940. doi:[10.3390/su13168940](https://doi.org/10.3390/su13168940).

Table 1. Studies that empirically explore ESG ratings divergence.¹

Authors	Year of the study	Data period	Asset 4 (now Refinitiv)	Beyond Ratings	Bloomberg	Calvert	DJSI	FTSE	FTSE4Good	GES	Innovest	Inrate	ISS Quality Score	KLD	MSCI	Vigeo-Eiris	Refinitiv	RepRisk	RibecoSAM	Sustainalytics	Thomson Reuters
Berg, F., Koebel, J.F., & Rigobon, R.	2022	2014 and 2017	x											x	x	x					
Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L.	2021	2020													x		x				
Bouyé, E., & Menville, D.	2021	2017		x											x			x			
Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S.	2016	2002-2010	x		x	x			x		x			x							
Christensen, D., Serafeim, G., & Sikochi, A.	2022	2004-2016	x												x						x
Dorflleitner, G., Halbritter, G., & Nguyen, M.	2015	2002-2012	x	x										x							
Gibson, R., Krueger, P., Riand, N., & Schmidt, P. S.	2021	2013-2017	x	x				x				x		x	x						x
Lopez, C., Contreras O., & Bendix, J.	2020	2018																		x	x
Romero, S., Jeffers, A. E., Aquilino, F., & DeGaetano, L.	2018	2016			x															x	x
Semenova, N., & Hassel, L.G.	2015	2003-2011	x							x				x							
Widyawati, L.	2020	2009-2013			x									x							x
Zumente, I. & Lacey, N.	2021	not stated			x								x	x						x	x

¹ There has been several mergers and acquisitions of rating institutes of the years, and they are here listed by the name that was current at the time of the study.