Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings

Dane Christensen
University of Oregon
danec@uoregon.edu

George Serafeim
Harvard University
gserafeim@hbs.edu

Anywhere Sikochi*
Harvard University
ssikochi@hbs.edu

November 2019

Abstract

Despite the rising use of environmental, social, and governance (ESG) ratings in financial markets, there is substantial disagreement across rating agencies regarding what rating to give to individual firms. As what drives this disagreement is unclear, we examine the extent to which a firm's ESG disclosure and average ESG rating explain this disagreement. Contrary to conventional wisdom that greater disclosure helps reduce disagreement, our findings suggest that greater ESG disclosure leads to greater disagreement across ESG rating agencies. These findings hold using firm fixed effects, changes models, and using a difference-in-differences design with staggered mandatory ESG disclosure shocks. We also find that rating disagreement is greater when firms have high or low average ESG ratings, relative to firms with medium average ESG ratings. Overall, our findings highlight the difficulty that firms face in resolving ESG rating disagreement and the need for developing rules and norms for evaluating ESG information.

Keywords: ESG Ratings; Rating Agency Disagreement; ESG Disclosure; Corporate Social Responsibility; Sustainability

JEL Classification: G24, M14, M41, Q56

We thank Vishal Baloria, Mark Bradshaw, Mary Ellen Carter, Jeff Doyle, Jacquelyn Gillette, Michelle Hanlon, Amy Hutton, Zach Kaplan, Brad Lindsey, Christopher Noe, Sugata Roychowdhury, Susan Shu, Chad Simon, Chris Skousen, Eric So, Rodrigo Verdi, Joe Weber, and workshop participants at Boston College, MIT, Utah State University, and Washington University in St. Louis for useful comments and suggestions. We also thank Carolyn Liu and James Zeitler for research assistance. Christensen is grateful for financial support from the Lundquist College of Business. Serafeim and Sikochi are grateful for financial support from the Division of Faculty Research and Development at Harvard Business School.

* Corresponding author: Anywhere (Siko) Sikochi, Harvard Business School, Soldiers Field Rd, 370B Morgan Hall, Boston, MA 02163; Telephone: (617) 496-3756; Email: ssikochi@hbs.edu.

1. Introduction

One of the biggest developments in financial markets, in recent years, has been the integration of environmental, social and governance (ESG) information in investment decisions. According to recent estimates, more than $30 trillion in assets under management are invested using sustainable strategies that apply ESG criteria in investment analysis and portfolio selection (GSIA 2018). Seeking to capture ESG information that thousands of publicly listed firms increasingly disclose, investor spending on ESG ratings from data providers (i.e., ESG rating agencies) increased from $200 million to $500 million between 2014 and 2018 (Gilbert 2019).

In line with the increased use of ESG ratings by market participants, a growing number of academic studies have examined the association of ESG ratings with variables of interest, including stock market performance, accounting performance, financial constraints, and governance characteristics (e.g., Cheng, Ioannou, and Serafeim 2014; Khan, Serafeim, and Yoon 2016; Hubbard, Christensen, and Graffin 2017).[1] However, recently academic research (Chatterji et al. 2016) and commentators in the media have questioned these ratings noting the high degree of disagreement among data providers. For example, one *Wall Street Journal* article noted, "Environmental, social and governance criteria are hard to define. When we measure how different ESG providers rate companies in the S&P 500, there's often little overlap. By contrast, when ratings agencies score those same companies for their creditworthiness, they are much more often in agreement" (Sindreu and Kent 2018). Similarly, regulators have also voiced concerns about the substantial disagreement in ESG ratings. Commissioner Peirce of the Securities and Exchange Commission (SEC) mentioned in a recent speech that "the different [ESG] ratings available can

[1] ESG ratings are evaluations of a company based on a comparative assessment of their quality, standard or performance on environmental, social or governance (ESG) issues (Brackley, Petroy, and Wong 2019).

vary so widely, and provide such bizarre results that it is difficult to see how they can effectively guide investment decisions."[2] These examples highlight the lack of consistency in assigning ESG ratings to companies across different providers.

This issue is of great importance because in the absence of agreement on what good ESG performance constitutes, market participants might be misled by ESG ratings. As an article in the *Financial Times* noted "Investors need to be clear about what the methodology they choose is actually measuring, and why. Otherwise ESG scoring risks creating a false sense of confidence among investors who don't really understand what lies behind the numbers — and therefore don't really understand what they're buying" (Allen 2018).

While several articles have highlighted the fact that data providers greatly disagree on how to rate a company's ESG performance, we have very little evidence on why providers disagree so much. Without understanding the reasons for this disagreement, it is difficult to understand not only what the potential remedies could be, but also the plausible consequences of this disagreement.[3] Our paper therefore, without taking a stance on whether the disagreement is undesirable or not, focuses on documenting key firm characteristics that help explain why providers disagree when assigning ESG ratings to a given firm.

To do this, we analyze data from three of the largest providers of ESG ratings to investors: MSCI, Thomson Reuters, and Sustainalytics. These rating agencies act as the main information intermediaries for ESG information in financial markets, similar to how credit rating agencies act as information intermediaries and gatekeepers in financial markets. Based on their ESG ratings, our analysis includes 30,700 firm-year observations across 5,637 unique firms between the years

[2] https://www.sec.gov/news/speech/speech-peirce-061819

[3] In a concurrent working paper, Kimbrough, Wang, and Wei (2019) find that a measure of ESG disagreement across two rating agencies is positively associated with higher return volatility, higher bid-ask spreads, and higher analyst forecast dispersion and negatively associated with the presence of a GRI report.

2004 and 2016.

While the number of firm characteristics one could examine is large, we focus on two key firm attributes that we hypothesize are likely to be of first-order importance in causing providers to disagree on ESG ratings. The first attribute we focus on is the extent of a firm's ESG disclosure, as theory suggests that disagreement arises due to different information sets and/or different interpretations of information (e.g., Cookson and Niessner 2019). We use a firm's ESG disclosure score calculated for each year from Bloomberg as a proxy for the firm's ESG disclosure practices. Conventional wisdom and a plethora of evidence in other settings, such as credit ratings and analyst forecasts, suggest that higher disclosure would be associated with lower disagreement (e.g., Lang and Lundholm 1996; Morgan 2002; Hope 2003), as greater disclosure reduces information asymmetry. We argue that in our setting, due to the subjective nature of ESG information, the opposite would be true, where higher disclosure would be associated with higher disagreement, as disclosure expands opportunities for different interpretations of information.

Specifically, in the absence of disclosure, data providers are more likely to agree because they use similar rules of thumb and imputation techniques. For example, they likely perceive the lack of disclosure on issues that are widespread and significant for a given industry (Khan, Serafeim, and Yoon 2016) as a bad attribute and thereby assign bad performance to the company. Similarly, for less significant issues where few companies in an industry disclose information, they likely perceive the lack of disclosure as a sign that the issue is relatively unimportant and therefore impute the company's performance to be the average performance of companies in the industry.

In contrast, for companies with higher levels of disclosure, data providers need to make a judgment about whether the disclosure means good or bad performance. For example, a company that discloses lost-time injury rates needs to be judged based on this disclosure. This gives rise to

a level of subjectivity that leads to higher levels of disagreement. We argue that this subjectivity increases as firms expand their disclosures. This prediction is consistent with arguments from the sociology literature, which theorizes that a plurality of evaluations is likely to occur in newly emerging fields where rules and norms for evaluation are less developed (Lamont 2012). Higher disclosure also increases the likelihood that the ESG rating agencies might be able to use different metrics to evaluate a firm's performance on the same issue and therefore lead to greater rating disagreement.[4] Moreover, as in financial markets, evaluators may disagree over which measures are more relevant to assessing ESG performance. Collectively, these arguments suggest that greater ESG disclosure would result in greater ESG rating disagreement.

Our analyses provide strong support for this prediction. First, we provide descriptive evidence that although ESG disclosure has increased dramatically in the last two decades (through voluntary and mandatory disclosure efforts), the level of ESG disagreement for a given firm has in fact increased over the same period. The results are consistent in the multivariate models that control for a number of other firm characteristics including firm size, profitability, analyst coverage, institutional ownership, industry membership, and valuation multiples. We estimate panel regressions with industry and year fixed effects and find a strong positive relationship between ESG disclosure and disagreement. We also estimate models with firm fixed effects and changes models and find similarly strong results of a positive association.

Moreover, consistent with our explanation that lack of consensus on what constitutes good performance and what metrics to use lead to a positive association between disclosure and disagreement, we find that the environmental and social pillars of ESG, rather than governance,

[4] Agencies may use different metrics because of different ideologies or, as noted in prior literature (Merton 1987; Hirshleifer and Teoh 2003; Bonsall and Miller 2017), rating agencies could simply focus on different factors because processing all the information contained in disclosures may be too costly.

primarily drive this result. While what constitutes good environmental or social performance is still a contested issue, good governance has been a topic of discussion for decades (e.g., Gompers, Ishii, and Metrick 2003). Although disagreements still exist, a common view of good governance has been codified in governance codes around the world and has been embraced by institutional investors and data providers.

The second firm characteristic we examine is the firm's ESG performance, as captured by the firm's average ESG rating across all three data providers. Ex ante it is less clear to us, compared with the effect of ESG disclosure, whether firms that receive higher or lower average ESG ratings would exhibit lower levels of disagreement among rating agencies. On the one hand, it might be easier for ESG rating agencies to agree when the company is clearly deficient in its ESG practices. As companies invest to improve their ESG performance and adopt policies and strategies, it might be harder for investors to agree on their importance, evaluating them through different lenses and therefore disagreeing on the score they assign. On the other hand, it might be easier to agree on companies that adopt more policies and exhibit better performance on outcome metrics, given that these companies are showing commitment to improving their ESG performance.

Our results support the latter interpretation. The average ESG rating exhibits a strong negative relationship with disagreement, suggesting that ESG rating agencies disagree less as ESG performance increases. These results hold in panel specifications with industry or firm fixed effects and in changes models. However, we also find that this is a rather simplistic view of the relationship between average ESG ratings and disagreement that hides significant non-linearities in the relationship. We document a significant U-shaped relationship, where the highest disagreement is among firms with the lowest ESG ratings, followed by firms with the highest ESG ratings. The lowest levels of disagreement are found among firms ranked in the middle of the

ratings distribution. These findings suggest that raters disagree a lot more when the stakes are high, as most investors are interested in identifying the best ESG performing companies in positive screening and best-in-class strategies, or the worst ESG performing companies in negative screening and engagement strategies (Amel-Zadeh and Serafeim 2018).

We also perform supplemental analyses to corroborate our primary inferences. First, to more rigorously address identification concerns, we use the staggered adoption of broad mandatory ESG disclosure requirements across countries as shocks to firm's ESG disclosures. Using a difference-in-differences design, we find that after a country or stock exchange implements mandatory ESG disclosure requirements, the affected firms tend to experience greater ESG rating disagreement (relative to the control firms). These findings provide greater confidence in our primary inferences.

Second, we explore the magnitude of the effect of ESG disclosures on ESG rating disagreement. Specifically, we investigate whether ESG disclosure ever leads to major rating disagreements, which are likely to be of the greatest interest to stakeholders that rely on ESG ratings. To examine this, we identify instances of extreme ESG disagreement by creating an indicator variable equal to one if ESG rating disagreement is in the top decile. We find that ESG disclosure also helps explain instances of extreme ESG rating disagreement. This suggests that ESG disclosure does indeed contribute to major disagreements about ESG performance.

Our study makes contributions to two streams of literature. First, we contribute to the literature that documents the presence of significant disagreement among ESG data providers (Chatterji et al. 2016). Our results suggest that this disagreement is most pronounced for firms with high levels of ESG disclosure and low average ESG performance ratings, thereby shedding light on the drivers of this disagreement. More broadly, our results contribute to the growing literature

that uses ESG data to understand their relationship with other important organizational and market outcomes (Khan, Serafeim, and Yoon 2016; Ferrell, Liang, and Renneboog 2016; Bereskin et al. 2018).

Second, we contribute to the literature that investigates rating disagreement in other settings, such as in credit ratings or analyst forecasts (e.g., Morgan 2002; Lang and Lundholm 1996; Bonsall and Miller 2017; Akins 2018). Contrary to evidence in these settings that disclosure mitigates disagreement, in our setting, disagreement is larger when firms have higher levels of disclosure. In our view this highlights the importance of developing a shared understanding of a) what constitutes good or bad ESG performance, and b) what metrics to use to capture ESG performance, as preconditions for transparency to decrease disagreement. It also highlights the challenge that firms currently face in using disclosure to mitigate ESG disagreement. Overall, given concerns over ESG rating disagreement, our findings suggest a lot of work still needs to be done to develop tools and norms to determine what characterizes good ESG performance.

2. Institutional Background and Hypothesis Development

2.1. ESG Rating Agencies

Several ESG data providers have emerged in the last two decades, most of which provide aggregate ratings of a firm's ESG performance. Responding to the need for collecting, interpreting, aggregating, and distributing ESG data, ESG rating agencies have become important information intermediaries in capital markets. The rating agencies share a common objective, which is to measure the ESG performance of a company (MSCI 2018; Thomson Reuters 2017; Sustainalytics 2018). The concept of ESG performance is intended to provide an assessment of how well a company is managing environmental, social, and governance risks and opportunities (MSCI 2018; Thomson Reuters 2017; Sustainalytics 2018). This performance is assessed by a wide array of

metrics that can be broadly classified into three groups: (1) policies and programs, (2) outcomes, and (3) controversies. Each of the rating providers cover very similar issues, although sometimes they label the issues with slightly different names. The exact metrics that rating agencies cover can differ, and how those metrics are used is a proprietary part of the rating process and is not observable to outside researchers.

Investors are the main audience of the ESG rating agencies featured in our study. The ESG ratings and the underlying data offered by the rating agencies are intended to help investors integrate ESG factors into their decisions, screen portfolios for risks and opportunities, generate investment ideas, conduct due diligence, determine opportunities for engagement, and support implementation of the UN PRI principles (MSCI 2018).

Of the ESG rating agencies, MSCI is widely considered the largest data provider to the investment community. It sells ESG ratings to investors, and also uses ESG data to construct stock market indices. Sustainalytics also sells its ESG ratings to investors, and provides other advisory and research services as well. Moreover, following its acquisition by Morningstar, Sustainalytics ratings form the basis for fund-level ESG ratings. Finally, ASSET4 was acquired by Thomson Reuters and its ESG ratings were integrated in Thomson Reuters' platforms and provided to their subscribers. Each of these rating agencies employs over 150 ESG analysts who collect and evaluate ESG data to produce ESG ratings. These ESG ratings, however, tend to differ greatly for a given firm across various rating agencies, which has drawn substantial criticism from outside observers.

2.2. ESG Rating Disagreement

It should be no surprise that ESG ratings would exhibit high levels of disagreement. A large body of work in sociology shows that human ability to make sense of information in a common way occurs over time and that both cultural and social processes define and enable the evaluation

of knowledge (Espeland and Stevens 1998; Knorr-Cetina 1999). In areas that are not highly formalized or that are newly emergent, it is more likely that pluralistic evaluative cultures prevail, leading to higher disagreement among raters (Lamont 2012), as in the case of the ESG field. In comparison to financial analysis that has been taught and formalized in the last century, ESG analysis has only emerged in the last two decades with most education and formalization activity taking place in the last ten years.

Moreover, because the information flow of ESG data is less formalized than the flow of financial data, which is systematized by institutional arrangements such as earnings calls and investor presentations, it can lead to further increases in disagreement. As analysts are more likely to receive and evaluate ESG information in a less structured way, analysts might seek and retrieve different pieces of information at different points in time and in a different sequence. This in turn leads to formation of differential expectations about a firm's ESG performance and therefore to different interpretations of subsequent information (e.g., Jones et al. 1968; Krüger and Nolte 2016). In other words, analysts' prior evaluations of specific ESG data can influence their subsequent evaluations of other ESG data, leading to heterogeneity in judgements (Lamont 2009). For example, a prior assessment of a firm's workplace practices might affect an analyst's evaluation of product safety.

In addition, evaluative practices that lead to ratings are influenced by conventions (Becker 1982). Whether evaluators follow customary rules is associated with how strongly they are invested in what defines a proper evaluation and, ultimately, their self-concept as an evaluator. For example, accountants, financial analysts, and credit analysts have established a strong identity through professional associations. No such identity yet exists for ESG analysts. Such a less institutionalized field will be less consistent in providing clear rules and in socializing new

evaluators, thereby giving rise to heterogeneity in judgements (Lamont 2012).

2.3. ESG Disclosure and Disagreement

With these things in mind, we now turn to the role that ESG disclosure may play in explaining ESG rating disagreement. Theory suggests that disagreement in financial markets generally arises due to individuals having different information sets and/or different interpretations of information (e.g., Cookson and Niessner 2019). Accordingly, ESG disclosure seems like a natural factor that could influence ESG rating disagreement. Existing literature has shown that disclosure improves a firm's information environment and resolves uncertainties about the firm for participants in both the equity and debt markets. In the equity markets, Lang and Lundholm (1996) provide evidence that a firm's disclosure practices lead to more accurate analyst earnings forecasts, less dispersion among individual analyst forecasts, and less volatility in forecast revisions. Moreover, their evidence suggests that disclosure reduces estimation risk and information asymmetry. Several subsequent studies corroborate these conclusions. For example, Hope (2003) finds that disclosure about a firm's accounting policies reduces uncertainty about its forecasted earnings and thus reduces forecast dispersion and forecast error. Overall, evidence suggests that disclosure reduces disagreement (i.e., forecast dispersion) among equity analysts.

Similarly, in the debt markets, disclosure has been found to reduce disagreement between credit rating analysts' assessment of credit risk of firms. Morgan (2002) finds that credit ratings agencies disagree more often over firms that are more opaque than other firms, and Bonsall and Miller (2017) find that firms with more readable disclosures are less likely to have split credit ratings. Additionally, Akins (2018) finds that firms with greater financial reporting quality are less likely to have split credit ratings. This evidence suggests that enhanced transparency through disclosures can mitigate rating disagreement on the premise that split ratings can arise when credit

rating agencies have limited access to data with which to base their credit assessments.

We argue that in the context of ESG ratings the opposite would be true, where more disclosure would be associated with higher disagreement. In contrast to financial disclosures where there is widespread agreement about the meaning of specific financial variables, such as leverage or profitability on the future creditworthiness of a company, for ESG disclosures there is no shared understanding yet on which exact metrics should be assessed to evaluate firm's ESG performance, or how to interpret and judge their meaning about a firm's ESG performance.

Anecdotal evidence suggests that this might be true. For example, when Workday Inc. significantly increased its ESG disclosure for fiscal year 2015 by adopting the new G4 Global Reporting Initiative guidelines,[5] its sustainability report increased from 54 to 98 pages. Its ESG disclosure score from Bloomberg increased significantly for both the environmental and social pillars. However, disagreement increased as its Thomson Reuter's rating for both environmental and social issues increased, MSCI's environmental rating decreased and its social rating increased, and Sustainalytics left its environmental rating unchanged while its social rating increased. We observed similar patterns when analyzing other companies that substantially increased their ESG disclosure, such as AT&T, Tech Mahindra and Altria, among others.

The rating disagreements can arise because more disclosure is more likely to lead to rating agencies using different metrics in assessing a firm's ESG performance. For example, on the issue of workplace safety, if a firm only discloses lost time injury rates, then all raters will likely use this metric to assess a company's performance. But if a firm also discloses additional information (e.g., the number of fatalities due to accidents, the number of lost workdays, time loss claims),

[5] G4 Global Reporting Initiative guidelines were issued by the Global Sustainability Standards Board (GSSB), an independent international organization whose core products are the Sustainability Reporting Standards.

raters might use different metrics to assess a firm's safety performance or might assign different importance to the different metrics.[6]

In addition, analysts might disagree on how to interpret a given metric. Specifically, in the absence of clear rules for evaluative practices, it is likely that rules of thumb will be developed for the least demanding evaluative tasks. For example, while an ESG analyst would need to judge a specific piece of information, in the absence of that information, a simple rule can be developed. Therefore, in the absence of disclosure, raters are more likely to agree because they use similar rules of thumb and imputation techniques. For instance, as some ESG analysts shared with us, if a single company does not disclose information on an important issue for its industry, they perceive the lack of disclosure as a bad attribute and thereby assign bad performance to the company. Similarly, in the absence of disclosure across most companies in an industry, they likely perceive the lack of disclosure as a sign that the issue is relatively unimportant, thereby imputing the company's performance to be the average performance of companies in the industry.

In contrast, for companies with higher levels of disclosure, ESG analysts need to make a judgement about whether the information being disclosed means good or bad performance and assess how to aggregate the different disclosures that a firm provides. For example, a company that discloses the presence of human rights policies or lost time injury rates needs to be judged based on these disclosures. This gives rise to a level of subjectivity that leads to ratings disagreement. We argue that this level of subjectivity increases as firms expand their disclosure, since firms first tend to disclose the presence or absence of policies, practices, and strategies (e.g., policies to provide employees with child care support or to require a health and safety policy) and then eventually disclose more specific quantitative metrics (e.g., employee turnover, CO_2

[6] For example, this can occur when raters have different ideologies about which measures are relevant to assessing ESG performance (e.g., Bloomfield and Fischer 2011).

emissions, water recycled). The latter are likely to be the elements that gives rise to the highest disagreement, as they need to be judged based on what constitutes "good" or "bad" performance, while at least in the context of policies, practices, and strategies, more disclosure might mean better performance. Collectively, all these arguments suggest that greater ESG disclosure increases ESG rating disagreement, because greater disclosure provides more information raters can disagree about, and thus creates more opportunities for raters to have different interpretations of information (Cookson and Niessner 2019). Therefore, our first hypothesis is as follows:

H1: ESG disclosure is positively associated with ESG rating disagreement.

2.4. ESG Performance and Disagreement

The literature has paid less attention to the issue of how firm ESG performance might be associated with disagreement. However, some evidence in other settings suggests that there is a relationship. For example, prior research has found that both the change in earnings relative to last-year and negative earnings are strongly positively associated with forecast dispersion, a result interpreted as showing that task complexity is higher for worse performing firms (Hope 2003). Closer to our paper, evidence in the credit rating literature suggests that the lower the average credit rating, the greater the disagreement among the agencies over bonds with low ratings than with high ratings (Cantor and Packer 1994). Subsequent analysis has documented non-linearities where the relationship between average credit ratings and disagreement is concave (Iannota 2006). Overall, the evidence has been interpreted in the literature suggesting that worse rated firms have higher uncertainty thereby leading to more disagreement.

It is less clear, in the ESG setting, whether firms that receive higher or lower average ESG ratings would exhibit lower disagreement. On the one hand, it might be easier for ESG rating agencies to agree when the company is clearly deficient in its ESG practices. As companies invest

to improve their ESG performance and adopt policies and strategies, it might be harder for investors to agree on their importance, evaluating them through different lenses and therefore disagreeing on the rating they assign. On the other hand, it might be easier to agree on companies that adopt more policies (e.g., hiring policy for diversity) and exhibit better performance on outcome metrics (e.g., a more diverse pool of internal promotions) given that these companies are showing a commitment to improving their ESG performance. For companies that show no commitment to ESG issues, it might be more difficult to distill why this might be happening and assess the implications for future ESG outcomes. This can give rise to task complexity and uncertainty effects, which have been documented in other settings. It is plausible that both effects are taking place, with the net effect being an empirical question. Therefore, our second hypothesis is non-directional:

H2: A firm's average ESG rating is not associated with the ESG rating disagreement.

3. Research Design and Sample Construction

3.1. Empirical Model

To test our hypotheses regarding how ESG rating disagreement is influenced by ESG disclosure (H1) and average ESG performance (H2), we estimate the following ordinary least squares (OLS) regression model:

$$ESG\_Disagreement_{it} = \beta_0 + \beta_1 * ESG\_Disclosure_{it} + \beta_2 * ESG\_Avg_{it} + \sum \beta_k * Controls + \varepsilon_{i,t} \quad (1)$$

where *ESG_Disagreement* is the standard deviation of ESG ratings firm *i* received from rating agencies for its performance in year *t*.[7] *ESG_Disclosure* is the firm's ESG disclosure score

[7] We opt to measure ESG disagreement using the standard deviation of ratings, as opposed to using the coefficient of variation, because we measure all the ESG ratings in our sample on the same scale (0 to 100) and scaling by a firm's

pertaining to year *t*'s performance. *ESG_Avg* is the average ESG rating a firm received from various rating agencies for year *t*'s performance. Note that each of these ESG variables is publicly released in year *t*+1, but pertain to year *t*'s performance, with a firm's ESG disclosure occurring before the ESG rating assessment. Thus the model allows for an assessment of how ESG disclosures influence subsequent ESG disagreement.

Controls consists of the following firm characteristics: firm size (*Firm Size*), firm performance (*ROA*), growth opportunities (*BTM*), capital structure (*Leverage*), the number of analysts following the firm (*Analyst Following*), and the percentage of shares held by institutional investors (*Inst. Ownership*). See Appendix A for further details on how each variable is measured. We also include industry, year, country, and rating agency fixed effects. For comparison, in some specifications we replace industry and country fixed effects with firm fixed effects to examine within-firm variation in ESG disagreement. We also run some specifications using a changes model, where we first-difference all the variables in the model. In all specifications, all continuous variables are winsorized at the 1st and 99th percentiles and standard errors are clustered by firm. Hypothesis 1 predicts that the estimated coefficient on *ESG_Disclosure* will be positively associated with *ESG_Disagreement* (i.e., $\beta_1 > 0$), and hypothesis 2 predicts that *ESG_Avg* will not be associated with *ESG_Disagreement* (i.e., $\beta_2 = 0$).

3.2. Sample Construction

To construct our sample, we obtained ESG ratings from three of the most prominent ESG rating agencies in the world. Specifically, we gathered ESG ratings from Morgan Stanley Capital

average rating would complicate inferences (Sørensen 2002) and induce a mechanical relationship between one of our hypothesized variables and our dependent variable. We also construct an alternative measure of ESG rating disagreement, measured as the average of absolute values of the difference between pairs of ratings that a firm receives for its performance in year *t*. Our inferences are generally similar using this alternative measure (untabulated.)

International's (MSCI) Intangible Value Assessment, Thomson Reuters' (TR) ASSET4, and Sustainalytics.[8] These ESG raters provide international coverage, which we later exploit to better get at identification. As some ESG rating agencies only provide one rating per firm each year, we construct a firm-year dataset by linking these ESG ratings to Worldscope. When multiple ESG ratings are released by a rating agency for a given firm-year's performance, we retain the last rating issued within 12 months of the firm's fiscal year-end, to ensure that all the rating agencies in our sample have had an opportunity to observe any ESG disclosures a firm has made relating to year *t*'s ESG performance.[9] To ensure that ratings from these agencies are comparable, we only use the ratings for the environmental, social, and governance pillars (as some raters also evaluate economic performance).[10] We re-scale the ratings when necessary so that they all range from 0 to 100. Specifically, while TR ASSSET4 and Sustainalytics scores range from 0 to 100, MSCI's scores range from 0 to 10. Thus, we multiply MSCI's scores by 10 to make them comparable.

To capture the extent of firms' ESG disclosure, we use ESG disclosure scores provided by Bloomberg. These scores ranges from 0.1 for firms that disclose a minimum amount of ESG data to 100 for firms that disclose every data point that Bloomberg collects. These disclosure scores are based on information firms disclose in various ways, such as via sustainability reports, annual reports, corporate websites, etc. Bloomberg's ESG disclosure coverage includes over 10,000 common stocks around the world and as a result is significantly larger (i.e., nearly double) than

[8] We do not include MSCI's KLD ratings in our analyses, as these ratings have been discontinued by MSCI and pertain almost exclusively to U.S. firms. However, our inferences are unchanged if we also include these ratings (untabulated). Some studies also use membership in the Dow Jones Sustainability Index, Calvert Index, and FTSE4Good Index as proxies for good ESG performance (e.g., Christensen 2016; Chatterji et al. 2016). These organizations do not publicly release ESG ratings, so we do not view them as rating agencies. Hence they are not included in our analyses.

[9] For example, if a firm has a December 31, 2010 year-end (i.e., fiscal year 2010), we use TR ASSET4's 2010 rating, and retain the last ESG rating MSCI and Sustainalytics released between January 1 and December 31, 2011.

[10] Because TR ASSET4's composite score also includes an economic component, we exclude this and calculate the overall ESG score as the average environmental, social, and governance score. Also note that while credit ratings agencies often use alphabetical ratings, historically few ESG rating agencies have translated their scores to alphabetical ratings.

the coverage universe of common stocks of the ESG rating agencies.

Our sample begins in 2004, the first year that we have data on the extent of firm's ESG disclosures, and ends in 2016 based on data availability at the time we constructed our sample. We obtain firm financial data from Worldscope, analyst coverage from the Institutional Brokers' Estimate System (I/B/E/S), and institutional ownership from Factset.

As we are interested in studying ESG rating disagreement, we require each firm-year to have ratings from more than one ESG rating agency. After requiring basic firm-level controls, along with ESG disclosure scores, our final sample consists of 30,700 firm-year observations from 5,637 unique firms. A summary of our sample construction process is outlined in Table 1.

4. Descriptive Analyses

4.1. Descriptive Statistics

Table 2 Panel A provides descriptive statistics illustrating the worldwide coverage of our sample. The sample is comprised of firms from 69 countries, with roughly three-quarters of the observations coming from nine countries: the United States, Japan, the United Kingdom, China, Canada, Australia, France, Germany, and Switzerland.

Table 2 Panel B provides descriptive statistics for the ESG ratings from each of the rating agencies in our sample. While our overall sample has 30,700 firm-year observations, each rating agency has fewer observations than this, as we only require each firm-year to be rated by at least two rating agencies, similar to research on analyst disagreement (e.g., Sadka and Scherbina 2007; Barinov 2013). In our sample, each of the raters tends to issue overall ESG scores of around 50 (out of 100 points possible). With that said, MSCI tends to issue slightly lower overall ESG ratings (48.4), and Sustainalytics tends to issue slightly higher overall ESG ratings (57.2). Additionally, ratings provided by TR tend to have the greatest variance across firms (i.e., standard deviation of

24.5), while ratings provided by Sustainalytics tend to have the least variance across firms (i.e., standard deviation of 9.8). Fairly similar trends exist for each of the underlying environmental, social, and governance pillars.

Table 2 Panel C provides descriptive statistics for ESG rating disagreement that exists for a given firm-year. While the average overall ESG rating a firm receives for a given year's performance is 52.8 (out of 100), there appears to be substantial disagreement across raters regarding what rating to give that firm. Specifically, the standard deviation of overall ESG ratings that a firm receives for a given firm-year is 12.3 (i.e., *ESG_Disagreement*). To help provide a more intuitive sense for the magnitude of this disagreement, we also document the absolute difference in ratings that pairs of agencies provide for a given firm-year. For example, we see that the ratings provided by TR and MSCI for a given firm-year differ by 19.7 points, on average. Similarly, the ratings TR and Sustainalytics provide for a given firm year differ by 15.8, on average. While the ratings from MSCI and Sustainalytics tend to differ by the least amount for a given firm-year (12.7), that difference is also quite large. Fairly similar patterns exist for each of the underlying environmental, social, and governance pillars as well. Overall, this descriptive evidence suggests that there is considerable disagreement across rating agencies regarding how to view a given firm's ESG performance.[11]

Table 2 Panel D provides descriptive statistics for the variables used in our regression analyses. Here we see that the average firm's ESG disclosure score is 28.5 (out of 100), with a reasonable amount of variation across firms (i.e., a standard deviation of 15.0) consistent with the

[11] In untabulated analyses, we also examine *ESG_Disagreement* by industry and find that there tends to be substantial disagreement across all Fama French 49 industries, with the tobacco industry having slightly higher disagreement than the other industries. If we examine *ESG_Disagreement* by country, we also find there tends to be substantial disagreement across almost all countries (untabulated). No particular country stands out as having much higher disagreement than the others. The few countries without much disagreement tend to have less than 10 observations (e.g., Kazakhstan, Ukraine, Kenya).

literature (e.g., Grewal, Hauptmann, and Serafeim 2017). We also see that the average firm in our sample is fairly large (total assets of US$ 7.3 billion), profitable (ROA of 0.05), and is followed by around 11 sell-side equity analysts.

The correlations in Table 3 also illustrate that ESG disagreement tends to be greater for larger firms, firms with better ESG disclosure, and firms with greater analyst following. Another interesting observation is that although *ESG_Disclosure* and *ESG_Avg* are positively correlated with each other, they have very different correlations with *ESG_Disagreement*.[12] Specifically, *ESG_Disclosure* is positively correlated with *ESG_Disagreement*, while *ESG_Avg* is negatively correlated with *ESG_Disagreement*.

4.2. Is ESG rating disagreement going down over time?

While the descriptive evidence presented thus far suggests that rating agencies tend to have strikingly different views on a given firm's ESG performance, it is possible that this disagreement has been declining over time and is not really much of an issue any more. This could occur if rating agencies' views begin to converge over time regarding what constitutes good ESG performance. This is plausible given efforts over time by such organizations as the GRI and SASB to create sustainability standards.[13] Thus, before moving onto the tests of our main hypotheses, we perform a few additional descriptive analyses.

First, we construct a time trend measure where we take the fiscal year for a particular firm-year observation and subtract the number 2004, which is the first year of our sample (*Time*). We then regress *ESG_Disagreement* on *Time* to examine whether ESG disagreement has been

[12] Variance inflation factors for all the variables in our models are below 3, indicating that multicollinearity is not a problem (Kennedy 2008).
[13] The Global Reporting Initiative (GRI) and Sustainability Accounting Standards Board (SASB) are independent non-profit organizations that develop sustainability reporting standards.

declining over time. As raters have been expanding their coverage over time (e.g., picking up smaller firms), we also include firm fixed effects in the model to ensure that changes in the sample composition over time don't confound our inferences. This effectively allows us to see how ESG disagreement for a given firm has been changing over time.

As reported in Table 4 Panel A we see that the coefficient on *Time* is positive and statistically significant, suggesting that ESG disagreement for a given firm has actually been increasing over time. This finding is inconsistent with rating agencies' views converging over time regarding what constitutes good ESG performance.[14] To illustrate how this trend compares to trends in ESG disclosure and average ESG performance ratings, we next replace the dependent variable with *ESG_Disclosure* and *ESG_Avg*, respectively, and re-run this analysis. In Table 4 Panel B we see that ESG disclosures also appear to be getting better over time (i.e., *Time* is again positive and statistically significant). And in Table 4 Panel C we see that average ESG ratings for a given firm have also been increasing over time.

Overall, these analyses suggest that despite improving ESG disclosures and ESG performance over time, ESG disagreement appears to be increasing over time. However, these analyses do not speak to whether these trends are actually related.

5. Empirical Results

5.1. What drives ESG rating disagreement?

5.1.1. Portfolio analyses of ESG rating disagreement

We now turn to examining our primary hypotheses regarding whether ESG disclosure and

[14] After our sample period, Thomson Reuters revised its ESG rating methodology. As they released ratings using the new methodology that also pertain to prior years, we re-ran this analysis using these new ratings as an input into the calculation of *ESG_Disagreement*. The results are very similar (untabulated). Results from our main tests of hypotheses are also very similar using these new ratings (untabulated).

ESG performance play a role in influencing ESG ratings disagreement. We begin by tabulating univariate statistics by allocating firms into portfolios. Specifically, we create 3x3 portfolios, based on a firm's level of ESG disclosure and average rating of ESG performance, our two key variables of interest. To create these portfolios, we rank *ESG_Disclosure* into terciles (high, medium, and low), and also rank *ESG_Avg* into terciles (high, medium, and low). This yields nine portfolios.

For each portfolio, we tabulate the mean and median ESG rating disagreement in Table 5 Panels A and B, respectively. A few key insights are immediately noticeable. First, across all the portfolios we see that disagreement increases monotonically as disclosure increases, consistent with hypothesis 1. Notably, the spread in disagreement across disclosure terciles is highest for the firms rated as the worst ESG performers. One way to interpret this finding is that rating agencies have a more difficult time agreeing on the interpretation of additional disclosures when firms have poor ESG performance.

Regarding hypothesis 2, it appears that rating agencies tend to disagree the most regarding firms with low average ratings. Both panels show that disagreement declines for firms with better ESG performance ratings. With that said, the relationship between ESG disagreement and average ESG ratings appears to be somewhat non-linear. There seems to be a U-shaped relationship where the lowest level of disagreement is found for the firms in the middle tercile. We revisit and investigate this non-linearity in our multivariate models (see Section 6.3).[15]

5.1.2. Multivariate tests of ESG rating disagreement

Given that these portfolio statistics do not account for a host of other factors such as

[15] An interesting observation is that the largest spread in disagreement between the high and low average ESG performance rating portfolios is observed for the high ESG disclosure portfolios. A plausible interpretation is that in the presence of significant disclosure, the lack of shared understanding of the meaning of the disclosures is more pronounced for firms that are clearly bad performers.

differences across industries and other firm characteristics we next turn to the multivariate analyses. Table 6 reports the results of our tests based on Equation 1 using three different specifications: (1) without firm fixed effects, (2) with firm fixed effects, and (3) a changes model.

Regarding our first hypothesis that ESG disclosure will be positively associated with ESG disagreement, we find that the estimated coefficient on *ESG_Disclosure* is indeed positive and statistically significant across all the models. These findings suggest that when a firm increases its ESG disclosure, this appears to exacerbate ESG disagreement across rating agencies. In other words, greater disclosure appears to give rating agencies simply more information to disagree about and interpret differently. In terms of the magnitude of this effect, using the model with firm fixed effects we observe that a one standard deviation increase in *ESG_Disclosure* is associated with a 1.41 increase in *ESG_Disagreement*. Given that the average *ESG_Disagreement* in the sample is 12.32, this appears to be an economically significant effect. Overall, these findings are consistent with H1.

Next, regarding our second hypothesis that ESG performance is unrelated to ESG disagreement, we find that the estimated coefficient on *ESG_Avg* is negative and statistically significant across all specifications. These findings suggest that ESG rating agencies tend to disagree more about firms with poor ESG performance.[16] In terms of economic significance, using coefficients from the model with firm fixed effects we observe that a one standard deviation increase in *ESG_Avg* is associated with a 2.86 decrease in *ESG_Disagreement*. Again, this appears to be an economically meaningful effect, relative to the sample average for *ESG_Disagreement*.

In terms of other factors that help explain ESG disagreement, we see that firm fixed effects appear to explain a sizable amount of the variation in ESG disagreement. Specifically, moving

[16] Note that the negative association between *ESG_Avg* and *ESG_Disagreement*, as well as the positive association between *ESG_Disclosure* and *ESG_Disagreement,* are consistent with the univariate correlations.

from the model without firm fixed effects, to the model with firm fixed effects, we see the explanatory power of the model improves dramatically (i.e., the adjusted R^2 goes from 0.118 to 0.516) suggesting that a lot of the disagreement about a firm's ESG performance is very persistent over time.

As for the other firm characteristics in the model, most do not yield consistently significant results.[17] The one exception is institutional ownership, which is negative and significant across all the specifications, which suggests that firms with greater institutional ownership tend to have lower ESG disagreement. However the economic magnitude of this finding is relatively small. A one standard deviation increase in institutional ownership is associated with a 0.37 decrease in ESG disagreement, based on the estimates from the model with firm fixed effects. The magnitude of this effect is roughly a quarter of the size of the effect we observe for *ESG_Disclosure*.

5.2. What dimensions of ESG disclosure contribute to ESG disagreement?

While our results thus far suggest that greater ESG disclosure is exacerbating ESG disagreement, it is likely that some attributes of the disclosures play a greater role than others in explaining ESG disagreement. In particular, disclosures relating to environmental and social issues seem more likely to generate disagreement, as what constitutes good performance in these areas is less well established, relative to what constitutes good governance.

Using the separate disclosure scores Bloomberg provides for each ESG pillar, we test which ESG pillars are primarily responsible for the positive association between ESG disclosure and ESG disagreement.[18] Specifically, we re-run our main analysis and replace *ESG_Disclosure*

[17] If we add the firm's annual market-adjusted stock return to the model, it is also not statistically significant. We exclude it from our primary model, as adding this variable reduces our sample size by several thousand observations.
[18] Although we would like to perform textual analyses on firms' ESG disclosures (e.g., sustainability reports) to investigate this issue, we are unfortunately unable to obtain the ESG disclosures made by the firms in our sample.

with its underlying component scores (*E_Disclosure*, *S_Disclosure*, and *G_Disclosure*). As reported in Table 7, we find that across all the specifications, *E_Disclosure* and *S_Disclosure* have a consistently positive and significant association with *ESG_Disagreement*. In contrast, in only one of the specifications (i.e., the OLS model without firm fixed effects) is *G_Disclosure* statistically significant). These findings are consistent with our expectations and suggest that firms' environmental and social disclosures appear to contribute to ESG rating disagreement, while firms' governance disclosures do not appear to generate as much controversy.

In Table 8, we corroborate this more directly by seeing whether disclosure for a particular pillar (e.g., environmental disclosure) helps explain disagreement for that pillar (e.g., environmental disagreement), while controlling for the average performance rating for that pillar (e.g., environmental performance). The results from these tests produce similar inferences, again suggesting that better environmental and social disclosures tend to contribute to disagreement, with governance disclosures not playing much of a role in disagreement.

6. Additional Analyses

6.1. Shocks to ESG Disclosure

One concern regarding our results is that they may somehow be influenced by self-selection or firm-level time-varying correlated omitted variables because firms can choose how much ESG information to disclose. Although we try to mitigate the possibility of such biases by estimating changes models and firm fixed effects models, it still could be the case that our results might be influenced by unobservable variables. To mitigate this concern, we exploit the passage of broad

While firms post their most recent sustainability reports on their websites, they generally do not post reports that are over a few years old. Thus getting access to historical reports is problematic. Although CorporateRegister.com has gathered a fairly comprehensive collection of sustainability reports over time, paid subscribers are only allowed to download up to 100 sustainability report pdfs a month. Given the large size of our sample, it is not feasible to gather the reports that pertain to our sample.

mandatory ESG disclosure requirements, which went into effect in numerous countries during our sample period. These broad ESG disclosure requirements arose due to legislation at the country level or listing requirements from individual stock exchanges. From an identification perspective, the attractive features of these mandatory disclosure requirements are that they are staggered across many countries over time, and they should lead to exogenous increases in ESG disclosure. A summary of the laws/regulations that were passed in different countries is shown in Appendix B.[19] Based on these laws/regulations, we create an indicator variable equal to one if the firm-year's ESG performance was subject to mandatory ESG disclosure requirements; zero otherwise (*Mandatory_Disclosure*).

We first confirm that these mandatory ESG disclosure requirements did indeed improve firms' ESG disclosures. Because the firms in our sample tend to be quite large, they may have already been voluntarily disclosing ESG information at the level required by these mandatory requirements. Thus it is possible that these mandatory disclosure requirements may not have resulted in an increase in ESG disclosure. Therefore we first perform a validation analysis before performing our difference-in-differences analysis. We do this by regressing *ESG_Disclosure* on *Mandatory_Disclosure*, with the same control variables as before, along with firm and year fixed effects. As shown in Table 9, the estimated coefficient on *Mandatory_Disclosure* is positive and statistically significant, confirming that these disclosure requirements improved firm's ESG disclosures, as expected.

We then perform a difference-in-differences analysis with firm and year fixed effects, where we estimate Equation 1 and replace *ESG_Disclosure* with *Mandatory_Disclosure*. As

[19] For further details on these disclosure requirements, see https://www.carrotsandsticks.net. Also note that our tests capture the first instance of broad ESG disclosure requirements in a country. Some countries have subsequently issued additional legislation that requires even greater ESG disclosure than the original mandate.

reported in Table 9, we observe that *Mandatory_Disclosure* is positive and statistically significant, consistent with our primary results. To address the parallel trends assumption in this analysis, we re-run this analysis and add an indicator variable equal to one in the year prior to when a firm became subject to mandatory ESG disclosure. As expected, this variable is statistically insignificant (untabulated), suggesting that the results documented here are not the result of a pre-existing trend.

These findings provide further support for the notion that greater ESG disclosure leads to greater ESG disagreement. That being said, a few caveats are worth highlighting. First, many of our observations that take the value of one for *Mandatory_Disclosure* come from a relatively small number of countries thereby making these results dependent on these countries. Second, some of the countries, despite having disclosure mandates, have historically had very low levels of disclosure suggesting that not all such mandates have been effective (e.g., Greece). With those caveats in mind, we believe that these tests need to be interpreted with caution, but are additive to the mosaic of evidence we present in the paper.

6.2. Extreme ESG Rating Disagreement

Although our results thus far indicate that ESG disclosures are contributing to greater disagreement among ESG rating agencies, it is still a bit unclear whether disclosure ever leads to major disagreements, which are likely to be of the greatest interest to stakeholders that rely on ESG ratings. To examine this, we identify instances of extreme ESG disagreement by creating an indicator variable equal to one if *ESG_Disagreement* is in the top decile; set to zero otherwise (*Extreme_Disagreement*). We then modify equation (1) by using *Extreme_Disagreement* as our dependent variable and re-run our main analysis. As reported in Table 10, we find that the estimated coefficient on *ESG_Disclosure* continues to be positive and statistically significant in

all specifications, suggesting that ESG disclosure contributes to extreme disagreements between rating agencies regarding a firm's ESG performance. Similarly, we see that *ESG_Avg* continues to be negative and statistically significant, suggesting that major disagreements are more likely to occur regarding firms that generally have poorer ESG performance.

6.3. Non-linearity tests

Although our multivariate results document a linear relationship between ESG disagreement and ESG disclosure and ESG performance, our univariate results suggests that a non-linear relationship may exist. Specifically, in the univariate analysis of the means across terciles of average ESG ratings (Table 5), we find a U-shaped relationship where the lowest level of disagreement is found for the firms in the middle tercile of average ESG ratings. To assess whether a non-linear relationship holds in a multivariate setting, we rank *ESG_Avg* into terciles and create an indicator variable equal to one if a firm-year is in the highest average ESG performance tercile; zero otherwise (*High ESG_Avg*). We also create an indicator variable equal to one if a firm-year is in the lowest average ESG performance tercile; zero otherwise (*Low ESG_Avg*). We then re-run Equation 1 where we replace *ESG_Avg* with these two indicator variables. This leaves firms with medium *ESG_Avg* as the comparison group. As reported in Table 11, we observe that both *High ESG_Avg* and *Low ESG_Avg* are positive and statistically significant. This suggests that rating agencies tend to disagree more for firms with high or low average ESG performance, relative to firms with medium average ESG performance. These findings are consistent with the U-shaped relationship noted in the portfolio analyses.

In untabulated analyses, we also examine whether there is a U-shaped relationship between *ESG_Disagreement* and *ESG_Disclosure*. Using an approach similar to the one we did for *ESG_Avg*, where we rank *ESG_Disclosure* into terciles and create two indicator variables, we do

not observe a U-shaped relationship. The relationship appears to be linear, where greater disclosure is associated with greater disagreement.

6.4. Pillar Weights

We have focused our analyses on the importance of disclosure and average ratings as explanatory variables. Another thing that could play a role in rating disagreement is disagreement on the weights raters place on the three ESG pillars. To test this, we created a weight disagreement measure for each firm-year by calculating the standard deviation of weights that raters used for each ESG pillar (e.g., environmental weight disagreement), and then took the average of the weight disagreement scores for the E, S, and G pillars (*ESG_Weight_Disagree*). As expected, the univariate correlation between this measure and ESG rating disagreement is positive and statistically significant ($p < 0.05$; untabulated). However, when we include this variable in our multivariate analyses, as reported in Table 12, *ESG_Weight_Disagree* is not statistically significant.[20] Thus, its influence on disagreement appears to be subsumed by the other variables in the model. Because most raters place different weights on the pillars by industry (e.g., MSCI 2018), with some deviations to account for idiosyncratic firm cases, it is not too surprising that our primary results are unaffected by the inclusion of this variable, as the model includes industry fixed effects. However, this analysis does not necessarily imply that weight disagreement is unimportant, as the importance of an industry-level effect could be subsumed in our models where the unit of analysis is the firm-year.

To shed more light on this issue, we move to an industry-year level analysis to understand whether weight disagreement explains rating disagreement across industries. We calculate all

[20] The results are similar if we instead just include the three pillar weight disagreement scores as separate variables in the model (instead of aggregating them into a single variable).

variables at the industry-year level by averaging the firm-year observations within each industry-year. The univariate correlation between rating disagreement and weight disagreement for the 576 industry-year observations is positive and statistically significant ($p < 0.01$; untabulated). Industries with highest weight disagreement are personal services, utilities, and chemicals. When we perform a multivariate analysis, which is also reported in Table 12, we see that the industry average weight disagreement is positively associated with the industry average ESG rating disagreement ($p < 0.05$). Thus, at the industry-year level, disagreement about what weights to assign each ESG pillar does help explain some of the disagreement in ESG ratings. Overall, the results suggest that weight disagreement does not affect our primary inferences about firm-level variation in disagreement, but could play a role in understanding industry-level variation in disagreement.

7. Conclusion

The use of environmental, social, and governance (ESG) ratings in financial markets has risen dramatically, and so has the disagreement across ESG rating agencies regarding what rating to give to an individual firm. The media has highlighted the existence of ESG rating disagreement, but there is a dearth of evidence on what drives this disagreement. In this paper, we document empirical evidence on the extent to which a firm's ESG disclosure and a firm's average ESG rating help explain this disagreement.

Our tests reveal several interesting findings. In contrast to evidence in other settings where greater disclosure helps reduce disagreement among information intermediaries, we find that greater ESG disclosure leads to greater ESG disagreement across ESG rating agencies. These findings appear to be driven by the environmental and social disclosures, rather than governance disclosures. Our results also show that there is a U-shaped relationship between a firm's average

ESG rating and ESG rating disagreement. Overall, we interpret our findings as evidence that rating agencies disagree more for firms that are performing very poorly or very well with respect to ESG, and that greater ESG disclosure does not help resolve such disagreement.

We note that over time as analysts develop a consensus both on the metrics to use to assess a firm's performance on a specific ESG issue and how to interpret the information reflected in each metric, the relation between disclosure and disagreement might diminish or even become negative. In other words, our study is likely to be reflective of the early stages of institutional innovation around nonfinancial disclosures.

We also note that it is not possible for us to separate the two mechanisms that we hypothesize contribute to the relation between disclosure and disagreement. To do so we would need to observe issues for which rating agencies are using the same metrics and then test whether, for those issues, firms that provide more disclosure have higher disagreement. Given that the rating agencies do not provide information on the exact metrics they use for all the issues they assess, we are not able to isolate those two mechanisms. Notwithstanding this caveat, our results shed light on why ratings are different, thereby advancing our understanding of ratings that increasingly represent important information for market participants.

As rating disagreements threaten to dampen confidence in ESG ratings, the findings in our study are important because without understanding why providers disagree, it becomes difficult to understand not only what the potential remedies could be, but also the plausible consequences of this disagreement.

References

Akins, B. 2018. Financial reporting quality and uncertainty about credit risk among ratings agencies. *The Accounting Review* 93(4): 1-22.

Allen, K. 2018. Lies, damned lies and ESG rating methodologies. *Financial Times*. Dec. 6. https://ftalphaville.ft.com/2018/12/06/1544076001000/Lies--damned-lies-and-ESG-rating-methodologies/

Amel-Zadeh, A. and Serafeim, G. 2018. Why and how investors use ESG information: Evidence from a global survey. *Financial Analysts Journal* 74(3): 87-103.

Barinov, A. 2013. Analyst disagreement and aggregate volatility risk. *Journal of Financial and Quantitative Analysis* 48(6): 1877-1900.

Becker H. 1982. Art Worlds. Berkeley: Univ. Calif. Press.

Bereskin, F., Byun, S.K., Officer, M.S. and Oh, J.M. 2018. The effect of cultural similarity on mergers and acquisitions: Evidence from corporate social responsibility. *Journal of Financial and Quantitative Analysis*, 53(5), pp.1995-2039.

Bloomfield, R. and Fischer, P.E., 2011. Disagreement and the Cost of Capital. *Journal of Accounting Research*, *49*(1), pp.41-68

Bonsall, S. and Miller, B. 2017. The impact of narrative disclosure readability on bond ratings and the cost of debt. *Review of Accounting Studies* 22(2): 608-643.

Brackley, A., Petroy, E., and Wong, C. 2019. Rate the Raters 2019: Expert Survey Results. SustainAbility, an ERM Group Company.

Cantor, R. and Packer, F. 1994. The credit rating industry. *Quarterly Review* 19(2): 1-26.

Chatterji, A., Durand, R., Levine, D. and Touboul, S. 2016. Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal* 37(8): 1597-1614.

Cheng, B., Ioannou, I. and Serafeim, G. 2014. Corporate social responsibility and access to finance. *Strategic Management Journal* 35(1): 1-23.

Christensen, D.M. 2016. Corporate accountability reporting and high-profile misconduct. *The Accounting Review* 91(2): 377-399.

Cookson, J.A, and M. Niessner. 2019. Why Don't We Agree? Evidence from a Social Network of Investors. *Journal of Finance*, forthcoming.

Espeland, W.N. and Stevens, M.L., 1998. Commensuration as a social process. *Annual Review of Sociology* 24(1): 313-343.

Ferrell, A., Liang, H. and Renneboog, L. 2016. Socially responsible firms. *Journal of Financial Economics* 122(3), pp.585-606.

Gilbert, M. 2019. The Rising Cost of Investing Responsibly. *Bloomberg*. Jan. 10. https://www.bloomberg.com/opinion/articles/2019-01-11/the-rising-cost-of-esg-and-socially-responsible-investing

Global Sustainable Investment Alliance (GSIA). 2018. *Global Sustainable Investment Review*. http://www.gsi-alliance.org/wp-content/uploads/2019/03/GSIR_Review2018.3.28.pdf.

Gompers, P., Ishii, J. and Metrick, A. 2003. Corporate governance and equity prices. *Quarterly Journal of Economics* 118(1): 107-156.

Grewal, J., Hauptmann, C. and Serafeim, G. 2017. Material sustainability information and stock price informativeness. HBS Working Paper.

Hirshleifer, D., and Teoh, S.H. 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* 36(1–3): 337–386.

Hope, O.K. 2003. Accounting policy disclosures and analysts' forecasts. *Contemporary Accounting Research* 20(2): 295-321.

Hubbard, T., Christensen, D. and Graffin, S. 2017. Higher highs and lower lows: The role of corporate social responsibility in CEO dismissal. *Strategic Management Journal* 38(11): 2255-2265.

Iannotta, G. 2006. Testing for opaqueness in the European banking industry: evidence from bond credit ratings. *Journal of Financial Services Research* 30(3): 287-309.

Jones, E., Rock, L., Shaver, K., Goethals, G. and Ward, L. 1968. Pattern of performance and ability attribution: An unexpected primacy effect. *Journal of Personality and Social Psychology* 10(4): 317-340.

Kennedy, P. 2008. A Guide to Econometrics. Sixth edition. Malden, MA: Blackwell Publishing.

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

Kimbrough, M., X. Wang and S. Wei. 2019. Do non-investor stakeholders benefit from voluntary disclosure? Evidence from corporate social responsibility reporting. Working paper.

Knorr-Cetina, K. 1999. Epistemic cultures: The cultures of knowledge societies. Cambridge, MA: Harvard.

Krüger, F. and Nolte, I. 2016. Disagreement versus uncertainty: Evidence from distribution forecasts. *Journal of Banking & Finance*, 72: S172-S186.

Lamont, M. 2009. *How professors think*. Harvard University Press.

Lamont, M. 2012. Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology* 38: 201-221.

Lang, M. and Lundholm, R. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* 71(4): 467-492.

Merton, R.C. 1987. A simple model of capital market equilibrium with incomplete information. *Journal of Finance* 42(3): 483–510.

Morgan, D.P. 2002. Rating banks: Risk and uncertainty in an opaque industry. *American Economic Review* 92(4): 874-888.

MSCI. 2018. MSCI ESG ratings methodology. Available at: https://www.msci.com

Sadka, R. and Scherbina, A. 2007. Analyst disagreement, mispricing, and liquidity. *Journal of Finance* 62(5): 2367-2403.

Sindreu, J. and Kent, S. 2018. Why It's So Hard to Be an 'Ethical' Investor. *The Wall Street Journal Online*. Sept. 1. https://www.wsj.com/articles/why-its-so-hard-to-be-an-ethical-investor-1535799601

Sørensen, J.B. 2002. The use and misuse of the coefficient of variation in organizational demography research. *Sociological Methods & Research* 30(4): 475-491.

Sustainalytics. 2018. Sustainalytics' ESG Rating Research Methodology. Available at: https://www.sustainalytics.com

Thomson Reuters. 2017. Thomson Reuters ESG Scores. March. Available at: https://www.thomsonreuters.com

APPENDIX A
Variable Definitions

Variable Name	Description	Source
ESG Variables		
ESG_Score	The firm's overall weighted environmental, social, and governance rating from a unique ESG rating agency for year *t* on a scale of 0 to 100.	MSCI IVA, TR Asset4, Sustainalytics
E_Score	The firm's environmental rating from a unique ESG rating agency for year *t* on a scale of 0 to 100.	MSCI IVA, TR Asset4, Sustainalytics
S_Score	The firm's social rating from a unique ESG rating agency for year *t* on a scale of 0 to 100.	MSCI IVA, TR Asset4, Sustainalytics
G_Score	The firm's governance rating from a unique ESG rating agency for year *t* on a scale of 0 to 100.	MSCI IVA, TR Asset4, Sustainalytics
ESG_Avg	The average environmental, social, and governance rating a firm received for year *t*'s ESG performance.	MSCI IVA, TR Asset4, Sustainalytics
E_Avg	The average environmental rating a firm received for year *t*'s environmental performance.	MSCI IVA, TR Asset4, Sustainalytics
S_Avg	The average social rating a firm received for year *t*'s social performance.	MSCI IVA, TR Asset4, Sustainalytics
G_Avg	The average governance rating a firm received for year *t*'s governance performance.	MSCI IVA, TR Asset4, Sustainalytics
ESG_Disagreement	The standard deviation of ESG ratings that a firm received for year *t*'s ESG performance.	MSCI IVA, TR Asset4, Sustainalytics
E_Disagreement	The standard deviation of environmental ratings that a firm received for year *t*'s environmental performance.	MSCI IVA, TR Asset4, Sustainalytics
S_Disagreement	The standard deviation of social ratings that a firm received for year *t*'s social performance.	MSCI IVA, TR Asset4, Sustainalytics
G_Disagreement	The standard deviation of governance ratings that a firm received for year *t*'s governance performance.	MSCI IVA, TR Asset4, Sustainalytics
Extreme_Disagreement	Indicator equal to 1 if *ESG_Disagreement* is in the top decile; set to zero otherwise.	

APPENDIX A (Continued)
Variable Definitions

Variable Name	Description	Source
Diff_TR_MSCI	The absolute value of the difference between the rating Thomson Reuters and MSCI gave a firm for year *t*'s ESG performance.	MSCI IVA, TR Asset4, Sustainalytics
Diff_TR_SUST	The absolute value of the difference between the rating Thomson Reuters and Sustainalytics gave a firm for year *t*'s ESG performance.	MSCI IVA, TR Asset4, Sustainalytics
Diff_MSCI_SUST	The absolute value of the difference between the rating MSCI and Sustainalytics gave a firm for year *t*'s ESG performance.	MSCI IVA, TR Asset4, Sustainalytics
ESG_Disclosure	The firm's ESG disclosure score for the ESG report pertaining to year *t*'s performance.	Bloomberg
E_Disclosure	The firm's environmental disclosure score for the ESG report pertaining to year *t*'s performance.	Bloomberg
S_Disclosure	The firm's social disclosure score for the ESG report pertaining to year *t*'s performance.	Bloomberg
G_Disclosure	The firm's governance disclosure score for the ESG report pertaining to year *t*'s performance.	Bloomberg
Mandatory_Disclosure	An indicator variable equal to one if the firm-year's ESG performance was subject to broad mandatory ESG disclosure requirements; zero otherwise.	Carrots & Sticks
ESG_Weight_Disagree	Calculated by first taking the standard deviation of weights that raters used for each ESG pillar (e.g., environmental weight disagreement), and then taking the average of the weight disagreement scores for the E, S, and G pillars.	MSCI IVA, TR Asset4, Sustainalytics
Control Variables		
Firm Size	The natural log of total assets (in millions of US$) as of the end of year *t*.	Worldscope
ROA	Return on Assets, defined as net income for year *t* divided by total assets at the end of year *t*.	Worldscope
BTM	The book value of equity divided by the market value of equity, as of the end of year *t*.	Worldscope
Leverage	Total liabilities divided by total assets, as of the end of year *t*.	Worldscope
Analyst Following	The natural log of the number of analysts following the firm as of the end of year *t*.	I/B/E/S
Inst. Ownership	Institutional Ownership, defined as the percentage of the firm's shares owned by institutional investors at the end of year *t,* multiplied by 100.	Factset

APPENDIX A (Continued)
Variable Definitions

Variable Name	Description	Source
Industry Fixed Effects	Industry indicators are based on the Fama French 49 industry classifications	Ken French's website
Time	The fiscal year for a particular firm-year observation minus the number 2004, which is the first year in our sample.	Worldscope

APPENDIX B
Mandatory ESG Disclosure Requirements (by Country)

Country	Source of Requirement	Year
Australia	Government	1996
Finland	Government	1997
France	Government	2003
Greece	Government	2007
Malaysia	Stock exchange	2007
Argentina	Government	2008
China	Stock exchange	2008
Sweden	Government	2008
Austria	Government	2009
Denmark	Government	2009
Portugal	Government	2009
Netherlands	Government	2010
Pakistan	Government	2010
South Africa	Stock exchange	2010
Philippines	Government	2011
India	Government	2012
Indonesia	Government	2012
Israel	Government	2012
Nigeria	Government	2012
Norway	Government	2013
Qatar	Government	2013
United Kingdom	Government	2013
Taiwan	Stock exchange	2014
Thailand	Government	2014
Turkey	Government	2014
Chile	Government	2015
Peru	Government	2015
Hong Kong	Stock exchange	2016

This appendix summarizes staggered adoptions of broad mandatory ESG disclosure requirements through the end of our sample period. Some ESG disclosure requirements are initiated by individual countries, and others are initiated by stock exchanges in those countries. The "year" column captures the first fiscal year-ends that became subject to the disclosure requirements. During our sample period, in a few countries the disclosure requirements only applied to a subset of firms. The requirements in Israel and Nigeria applied to banks, and in Qatar they applied to oil, energy, and transportation firms. In India they applied to firms on the BSE and National India stock exchanges, and in Australia they applied to firms on the Australian stock exchange. In Sweden they applied to state owned enterprises.

TABLE 1
Sample Construction

	Firm-Years	Unique Firms
Worldscope firm-years with ESG ratings (2004 - 2016)	65,995	11,217
Require ESG ratings from 2+ raters for each firm-year	(26,892)	(4,268)
	39,103	6,949
Require control variables	(511)	(91)
	38,592	6,858
Require ESG disclosure scores	(7,892)	(1,221)
	30,700	5,637

The table presents the summary of our sample construction. The sample covers the period from 2004 to 2016 for firm-years with at least two ESG ratings from MSCI, Thomson Reuters ASSET4, and Sustainalytics. Control variables and ESG disclosure scores are as described in Appendix A.

TABLE 2
Descriptive Statistics

Panel A: Country Composition

Nation	Freq.	%	Nation	Freq.	%
United States	9,704	31.61	Belgium	172	0.56
Japan	3,444	11.22	Turkey	165	0.54
United Kingdom	2,413	7.86	Bermuda	163	0.53
China	2,187	7.12	Philippines	147	0.48
Canada	1,848	6.02	Poland	136	0.44
Australia	1,561	5.08	Austria	134	0.44
France	852	2.78	Chile	130	0.42
Germany	728	2.37	Israel	97	0.32
Switzerland	561	1.83	New Zealand	97	0.32
India	534	1.74	Greece	90	0.29
South Africa	516	1.68	Portugal	81	0.26
Brazil	481	1.57	Colombia	77	0.25
Sweden	470	1.53	United Arab Emir.	58	0.19
Spain	362	1.18	Luxembourg	56	0.18
Italy	339	1.10	Jersey	46	0.15
Netherlands	332	1.08	Qatar	38	0.12
Singapore	316	1.03	Hungary	28	0.09
Malaysia	266	0.87	Czech Republic	25	0.08
Indonesia	222	0.72	Macau	21	0.07
Finland	218	0.71	Peru	21	0.07
Norway	217	0.71	Egypt	19	0.06
Ireland	215	0.70	Puerto Rico	12	0.04
South Korea	215	0.70	Cayman Islands	11	0.04
Russia	206	0.67	Malta	9	0.03
Mexico	202	0.66	Papua New Guinea	9	0.03
Denmark	198	0.64	Isle of Man	8	0.03
Thailand	187	0.61	Other	56	0.18
				30,700	100.00

TABLE 2 (continued)

Panel B: ESG Ratings from each Rating Agency

	N	Mean	Median	Standard Deviation
MSCI IVA				
ESG Score	26,116	48.36	48.00	12.41
E Score	26,108	50.20	50.00	21.28
S_Score	26,116	46.35	46.00	17.87
G Score	26,102	56.13	55.00	20.58
TR ASSET4				
ESG Score	29,767	54.13	56.15	24.53
E_Score	29,767	53.67	56.85	32.04
S Score	29,767	54.18	57.21	31.03
G Score	29,767	54.52	60.92	29.58
SUSTAINALYTICS				
ESG_Score	25,786	57.21	56.00	9.82
E Score	25,786	54.28	53.00	13.82
S Score	25,786	56.99	56.00	11.06
G Score	25,786	62.21	62.00	10.89

TABLE 2 (continued)

Panel C: Disagreement for each ESG category

	N	Mean	Median	Standard Deviation
ESG Overall				
ESG_Avg	30,700	52.80	52.95	13.93
ESG_Disagreement	30,700	12.32	11.65	6.75
ESG_Diff_TR_MSCI	25,183	19.67	17.90	13.01
ESG_Diff_TR_SUST	24,853	15.75	14.34	10.61
ESG_Diff_MSCI_SUST	21,202	12.71	11.00	9.08
Environmental				
E_Avg	30,700	52.08	52.23	19.11
E_Disagreement	30,700	17.03	16.60	8.38
E_Diff_TR_MSCI	25,175	27.47	26.05	17.29
E_Diff_TR_SUST	24,853	21.37	21.92	11.44
E_Diff_MSCI_SUST	21,194	17.28	15.00	12.28
Social				
S_Avg	30,700	52.08	52.34	16.90
S_Disagreement	30,700	17.45	17.22	8.52
S_Diff_TR_MSCI	25,183	27.10	25.49	17.09
S_Diff_TR_SUST	24,853	22.64	22.05	12.93
S_Diff_MSCI_SUST	21,202	18.06	15.67	13.16
Governance				
G_Avg	30,700	57.12	58.78	15.92
G_Disagreement	30,696	17.53	16.64	9.68
G_Diff_TR_MSCI	25,169	26.74	24.16	18.14
G_Diff_TR_SUST	24,853	23.46	20.62	15.89
G_Diff_MSCI_SUST	21,192	18.59	16.00	13.23

TABLE 2 (continued)

Panel D: Descriptive Statistics of Regression Variables (n = 30,700)

Variables	Mean	Standard Deviation	25%	Median	75%
ESG_Disagreement	12.32	6.75	7.12	11.65	16.92
E_Disagreement	17.03	8.38	11.14	16.60	22.33
S_Disagreement	17.45	8.52	11.19	17.22	23.12
G_Disagreement	17.53	9.68	10.06	16.64	24.09
ESG_Avg	52.80	13.93	41.43	52.95	64.01
E_Avg	52.08	19.11	35.71	52.23	68.36
S_Avg	52.08	16.90	38.09	52.34	65.90
G_Avg	57.12	15.92	45.11	58.78	69.44
ESG_Disclosure	28.50	14.97	14.91	25.62	40.50
E_Disclosure	19.20	18.79	0.00	13.95	34.88
S_Disclosure	26.00	19.32	8.77	23.33	38.60
G_Disclosure	51.78	9.03	48.21	51.79	57.14
Firm Size	8.89	1.64	7.77	8.73	9.87
ROA	0.04	0.08	0.01	0.04	0.08
BTM	0.64	0.50	0.30	0.52	0.85
Leverage	0.58	0.23	0.43	0.58	0.74
Analyst Following	2.43	0.73	2.08	2.56	2.94
Inst. Ownership	48.16	31.89	19.39	42.00	78.44

The table presents the sample descriptive statistics. Panel A shows the sample distribution by country (China includes observations from Taiwan and Hong Kong). Panel B reports summary statistics of ESG ratings from each rating agency. Panel C reports summary statistics of ESG rating disagreement for each ESG category. Panel D reports summary statistics of the control variables. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions.

TABLE 3

Pearson Correlations

		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
1.	*ESG_Disagreement*																	
2.	*E_Disagreement*	**0.36**																
3.	*S_Disagreement*	**0.53**	**0.15**															
4.	*G_Disagreement*	**0.25**	**0.02**	**0.08**														
5.	*ESG_Avg*	**-0.05**	**-0.05**	**-0.04**	**-0.12**													
6.	*E_Avg*	-0.01	0.00	**0.02**	**0.06**	**0.86**												
7.	*S_Avg*	**-0.05**	**-0.03**	**-0.07**	**-0.02**	**0.90**	**0.75**											
8.	*G_Avg*	**-0.04**	**-0.04**	**-0.07**	**-0.39**	**0.60**	**0.28**	**0.36**										
9.	*ESG_Disclosure*	**0.12**	0.00	**0.11**	**0.10**	**0.70**	**0.71**	**0.68**	**0.23**									
10.	*E_Disclosure*	**0.12**	0.00	**0.12**	**0.15**	**0.65**	**0.70**	**0.64**	**0.14**	**0.97**								
11.	*S_Disclosure*	**0.09**	-0.01	**0.09**	**0.06**	**0.65**	**0.62**	**0.66**	**0.23**	**0.90**	**0.79**							
12.	*G_Disclosure*	**0.09**	**0.03**	**0.06**	**-0.14**	**0.50**	**0.38**	**0.39**	**0.47**	**0.60**	**0.46**	**0.49**						
13.	*Firm Size*	**0.11**	0.01	**0.07**	**0.07**	**0.33**	**0.35**	**0.35**	**0.03**	**0.38**	**0.38**	**0.30**	**0.24**					
14.	*ROA*	**0.02**	**-0.03**	-0.01	0.01	**0.03**	**0.03**	**0.04**	**0.04**	0.01	0.00	**0.02**	**0.02**	**-0.16**				
15.	*BTM*	0.00	**-0.02**	**0.02**	**0.09**	**-0.02**	0.01	**0.03**	**-0.15**	**0.08**	**0.10**	**0.07**	**-0.04**	**0.28**	**-0.35**			
16.	*Leverage*	**0.02**	**0.02**	0.00	**-0.06**	**0.12**	**0.08**	**0.13**	**0.07**	**0.08**	**0.06**	**0.07**	**0.09**	**0.48**	**-0.25**	**-0.02**		
17.	*Analyst Following*	**0.02**	0.01	**-0.02**	**-0.03**	**0.38**	**0.35**	**0.36**	**0.21**	**0.32**	**0.30**	**0.29**	**0.26**	**0.32**	**0.13**	**-0.13**	**0.06**	
18.	*Inst. Ownership*	**-0.05**	**0.04**	**-0.08**	**-0.25**	**0.04**	**-0.12**	**-0.13**	**0.43**	**-0.20**	**-0.24**	**-0.23**	**0.21**	**-0.15**	**0.03**	**-0.24**	**0.02**	**0.09**

This table presents correlations for the variables in our tests (N=30,700). Bolded correlations are significant at the 0.05 level or lower. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions.

TABLE 4
Time Trend Analyses

Panel A: ESG rating disagreement over time

Dependent Variable:	ESG_Disagreement		
	Coef.	t-stat.	
Time	0.121	5.98	***
Intercept	11.283	65.12	***
Firm F.E.	Yes		
Adj. R^2	0.486		
N	30,700		

Panel B: ESG disclosure over time

Dependent Variable:	ESG_Disclosure		
	Coef.	t-stat.	
Time	1.388	51.89	***
Intercept	16.609	72.45	***
Firm F.E.	Yes		
Adj. R^2	0.870		
N	30,700		

Panel C: Average ESG performance over time

Dependent Variable:	ESG_Avg		
	Coef.	t-stat.	
Time	0.764	29.85	***
Intercept	46.256	210.9	***
Firm F.E.	Yes		
Adj. R^2	0.872		
N	30,700		

This table reports the trends over time of ESG rating disagreement in Panel A, ESG disclosure in Panel B, and Average ESG performance in Panel C. Time is the difference between the fiscal-year for a particular firm-year observation and 2004, which is the first year in the sample. All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 5
Portfolio Analyses

Panel A: Mean ESG Disagreement

		ESG Avg			
		Low	Medium	High	Diff
ESG_Disclosure	Low	12.88	8.92	11.43	-1.45
	Medium	14.23	9.83	12.72	-1.51
	High	17.51	12.03	13.69	-3.82
	Diff	4.63	3.11	2.26	

Panel B: Median ESG Disagreement

		ESG Avg			
		Low	Medium	High	Diff
ESG_Disclosure	Low	12.31	7.87	10.53	-1.78
	Medium	13.32	8.76	12.73	-0.59
	High	17.88	10.91	13.55	-4.33
	Diff	5.57	3.04	3.02	

This table presents univariate results from portfolio analyses of ESG rating disagreement. For each year we allocate each firms into terciles of ESG disclosure and terciles of the average ESG score across the rating agencies. Panel A (B) shows the mean (median) ESG disagreement across ratings for each of the nine portfolios of companies. Difference is calculated as High minus Low.

TABLE 6
Is better ESG Disclosure or Performance associated with greater ESG Rating Disagreement?

Dependent Variable: ESG Disagreement

	OLS			OLS w/ Firm F.E.			Changes Model		
	Coef.	t-stat.		Coef.	t-stat.		Coef.	t-stat.	
ESG Disclosure	**0.127**	**16.29**	***	**0.094**	**10.23**	***	**0.053**	**5.99**	***
ESG Avg	**-0.158**	**-19.32**	***	**-0.206**	**-18.09**	***	**-0.214**	**-18.22**	***
Firm Size	0.927	12.55	***	-0.015	-0.08		-0.157	-0.74	
ROA	1.695	2.20	**	-0.508	-0.65		0.717	1.16	
BTM	-0.437	-2.86	***	-0.287	-1.96	*	-0.012	-0.10	
Leverage	-0.146	-0.36		1.719	2.84	***	0.854	1.32	
Analyst Following	-0.410	-3.34	***	0.073	0.53		-0.045	-0.36	
Inst. Ownership	-0.010	-2.75	***	-0.012	-1.94	*	-0.017	-2.89	***
Intercept	14.848	3.66	***	13.820	4.16	***	-0.115	-0.26	
ESG Rater F.E.	Yes			Yes			Yes		
Year F.E.	Yes			Yes			Yes		
Industry F.E.	Yes			No			Yes		
Country F.E.	Yes			No			Yes		
Firm F.E.	No			Yes			No		
Adj. R^2	0.118			0.516			0.053		
N	30,700			30,700			24,234		

This table reports the results of tests based on Equation 1 examining the effects of ESG disclosure and Average ESG performance on ESG Disagreement. There are three specifications: (1) ordinary least squares (OLS) without firm fixed effects, (2) OLS with firm fixed effects, and (3) a changes model. All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 7
Is better ESG Disclosure for each ESG pillar associated with greater ESG Rating Disagreement?

Dependent Variable: ESG Disagreement

	OLS			OLS w/ Firm F.E.			Changes Model		
	Coef.	t-stat.		Coef.	t-stat.		Coef.	t-stat.	
E Disclosure	**0.075**	**10.16**	***	**0.047**	**5.95**	***	**0.031**	**4.37**	***
S Disclosure	**0.018**	**2.82**	***	**0.036**	**5.20**	***	**0.012**	**1.93**	*
G Disclosure	**0.039**	**3.77**	***	-0.003	-0.34		0.004	0.45	
ESG Avg	-0.158	-19.29	***	-0.207	-18.15	***	-0.214	-18.20	***
Firm Size	0.913	12.32	***	-0.034	-0.18		-0.154	-0.72	
ROA	1.732	2.25	**	-0.493	-0.63		0.710	1.15	
BTM	-0.437	-2.86	***	-0.281	-1.91	*	-0.014	-0.12	
Leverage	-0.119	-0.29		1.692	2.80	***	0.848	1.31	
Analyst Following	-0.406	-3.30	***	0.075	0.54		-0.043	-0.35	
Inst. Ownership	-0.010	-2.75	***	-0.011	-1.93	*	-0.017	-2.88	***
Intercept	14.744	3.63	***	15.015	4.49	***	-0.120	-0.27	
ESG Rater F.E.	Yes			Yes			Yes		
Year F.E.	Yes			Yes			Yes		
Industry F.E.	Yes			No			Yes		
Country F.E.	Yes			No			Yes		
Firm F.E.	No			Yes			No		
Adj. R^2	0.119			0.517			0.052		
N	30,700			30,700			24,234		

This table reports the results of tests examining the effects of ESG disclosure for each ESG pillar on ESG rating Disagreement. There are three specifications: (1) ordinary least squares (OLS) without firm fixed effects, (2) OLS with firm fixed effects, and (3) a changes model. All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 8
Is better Pillar-specific Disclosure associated with greater Pillar-specific Rating Disagreement?

Panel A: Environmental Rating Disagreement

Dependent Variable: E Disagreement

	OLS			**OLS w/ Firm F.E.**			**Changes Model**		
	Coef.	t-stat.		Coef.	t-stat.		Coef.	t-stat.	
E Disclosure	0.008	1.14		**0.026**	**2.90**	***	**0.031**	**3.41**	***
E Avg	0.006	0.99		-0.026	-2.49	**	-0.014	-1.17	
Intercept	11.269	1.96	*	13.620	3.16	***	-0.641	-1.10	
Firm Controls	Yes			Yes			Yes		
ESG Rater F.E.	Yes			Yes			Yes		
Year F.E.	Yes			Yes			Yes		
Industry F.E.	Yes			No			Yes		
Country F.E.	Yes			No			Yes		
Firm F.E.	No			Yes			No		
Adj. R^2	0.066			0.432			0.007		
N	30,700			30,700			24,234		

Panel B: Social Rating Disagreement

Dependent Variable: S Disagreement

	OLS			**OLS w/ Firm F.E.**			**Changes Model**		
	Coef.	t-stat.		Coef.	t-stat.		Coef.	t-stat.	
S Disclosure	**0.080**	**12.53**	***	**0.058**	**7.22**	***	**0.026**	**3.29**	***
S_Avg	-0.124	-17.56	***	-0.229	-22.25	***	-0.207	-18.38	***
Intercept	17.493	3.55	***	18.980	5.07	***	-1.295	-2.48	**
Firm Controls	Yes			Yes			Yes		
ESG Rater F.E.	Yes			Yes			Yes		
Year F.E.	Yes			Yes			Yes		
Industry F.E.	Yes			No			Yes		
Country F.E.	Yes			No			Yes		
Firm F.E.	No			Yes			No		
Adj. R^2	0.082			0.435			0.045		
N	30,700			30,700			24,234		

TABLE 8 (Continued)

Panel C: Governance Rating Disagreement

Dependent Variable: G Disagreement

	OLS			OLS w/ Firm F.E.			Changes Model		
	Coef.	t-stat.		Coef.	t-stat.		Coef.	t-stat.	
G Disclosure	**0.036**	**3.25**	***	-0.013	-1.06		-0.004	-0.25	
G Avg	-0.166	-24.26	***	-0.144	-14.58	***	-0.118	-10.42	***
Intercept	28.025	4.17	***	23.989	5.34	***	-1.115	-1.65	*
Firm Controls	Yes			Yes			Yes		
ESG Rater F.E.	Yes			Yes			Yes		
Year F.E.	Yes			Yes			Yes		
Industry F.E.	Yes			No			Yes		
Country F.E.	Yes			No			Yes		
Firm F.E.	No			Yes			No		
Adj. R^2	0.282			0.503			0.030		
N	30,696			30,696			24,230		

This table reports the results of tests examining the effects of ESG disclosure for each ESG pillar on disagreement for each respective pillar. Panel A reports the Environment pillar (E), Panel B the Social pillar (S), and Panel C the Governance pillar (G). There are three specifications: (1) ordinary least squares (OLS) without firm fixed effects, (2) OLS with firm fixed effects, and (3) a changes model. All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 9
Mandatory ESG Disclosure

	Validation			Diff-in-Diff		
Dependent Variable:	ESG Disclosure			ESG Disagreement		
	Coef.	t-stat.		Coef.	t-stat.	
Mandatory Disclosure	**1.431**	**5.26**	***	**0.432**	**2.14**	**
ESG Avg	0.273	21.42	***	-0.181	-15.81	***
Firm Size	0.756	3.32	***	0.051	0.26	
ROA	0.196	0.23		-0.473	-0.6	
BTM	0.242	1.28		-0.271	-1.83	*
Leverage	0.091	0.12		1.760	2.85	***
Analyst Following	0.728	4.35	***	0.145	1.06	
Inst. Ownership	-0.005	-0.63		-0.012	-1.97	**
Intercept	-17.617	-3.56	***	12.153	3.57	***
ESG Rater F.E.	Yes			Yes		
Year F.E.	Yes			Yes		
Firm F.E.	Yes			Yes		
Adj. R^2	0.885			0.511		
N	30,700			30,700		

This table reports the results using mandatory ESG disclosure requirements, which went into effect in different countries and at different times. The validation test confirms that mandatory disclosure requirements improve firms' ESG disclosures. The difference-in-difference (Diff-in-Diff) shows the difference-in-difference analysis with firm and year fixed effects. All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 10
Is better ESG Disclosure associated with Extreme ESG Rating Disagreement?

Dependent Variable: Extreme Disagreement

	OLS			OLS w/ Firm F.E.		
	Coef.	t-stat.		Coef.	t-stat.	
ESG Disclosure	**0.004**	**10.50**	***	**0.002**	**4.12**	***
ESG Avg	**-0.005**	**-14.47**	***	**-0.007**	**-13.22**	***
Firm Size	0.033	9.79	***	-0.006	-0.74	
ROA	0.036	1.08		-0.108	-2.81	***
BTM	-0.014	-2.13	**	0.000	0.05	
Leverage	0.003	0.18		0.093	3.36	***
Analyst Following	-0.014	-2.49	**	0.001	0.20	
Inst. Ownership	0.000	-3.05	***	0.000	-0.80	
Intercept	0.031	0.18		0.219	1.33	
ESG Rater F.E.	Yes			Yes		
Year F.E.	Yes			Yes		
Industry F.E.	Yes			No		
Country F.E.	Yes			No		
Firm F.E.	No			Yes		
Adj. R^2	0.071			0.373		
N	30,700			30,700		

This table documents whether ESG disclosure is associated with major ESG rating disagreement. The dependent variable, Extreme_Disagreement, is an indicator variable equal to 1 if ESG disclosure is in the top decile for each fiscal-year, and zero otherwise. There are two specifications: (1) ordinary least squares (OLS) without firm fixed effects, and (2) OLS with firm fixed effects. All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 11
Is there a non-linear relationship between ESG Performance and ESG Rating Disagreement?

Dependent Variable: ESG Disagreement

	OLS			**OLS w/ Firm F.E.**		
	Coef.	t-stat.		Coef.	t-stat.	
High ESG Avg	**1.752**	**9.66**	***	**0.376**	**2.08**	**
Low ESG Avg	**4.135**	**24.50**	***	**2.665**	**14.81**	***
ESG Disclosure	0.066	9.19	***	0.059	6.51	***
Firm Size	0.745	10.35	***	-0.202	-1.06	
ROA	1.879	2.50	*	-0.279	-0.36	
BTM	-0.253	-1.69		-0.264	-1.76	*
Leverage	0.025	0.06		1.812	2.95	***
Analyst Following	-0.426	-3.58	***	0.020	0.15	
Inst. Ownership	-0.011	-3.10	***	-0.014	-2.27	**
Intercept	7.139	1.90	*	3.162	1.05	
ESG Rater F.E.	Yes			Yes		
Year F.E.	Yes			Yes		
Industry F.E.	Yes			No		
Country F.E.	Yes			No		
Firm F.E.	No			Yes		
Adj. R^2	0.133			0.506		
N	30,700			30,700		

This table documents whether a non-linear relationship exists between Average ESG performance and ESG rating disagreement. There are two specifications: (1) ordinary least squares (OLS) without firm fixed effects, and (2) OLS with firm fixed effects. All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

TABLE 12
Multivariate association between ESG Rating Disagreement and ESG Weight Disagreement

Dependent Variable: ESG Disagreement

	Firm-Year Test			Industry-Year Test		
	Coef.	t-stat.		Coef.	t-stat.	
ESG Weight Disagree	2.090	1.28		**17.949**	**2.10**	**
ESG Disclosure	0.127	16.30	***	0.141	1.91	*
ESG Avg	-0.158	-19.32	***	-0.167	-2.50	**
Firm Size	0.932	12.60	***	0.085	0.22	
ROA	1.685	2.19	**	13.610	1.05	
BTM	-0.439	-2.87	***	0.368	0.29	
Leverage	-0.158	-0.39		-0.651	-0.21	
Analyst Following	-0.412	-3.35	***	0.602	0.91	
Inst. Ownership	-0.010	-2.73	***	0.002	0.08	
Intercept	11.867	7.88	***	13.065	3.92	***
ESG Rater F.E.	Yes			No		
Year F.E.	Yes			Yes		
Industry F.E.	Yes			No		
Country F.E.	Yes			No		
Firm F.E.	No			No		
Adj. R^2	0.119			0.131		
N	30,700			576		
Unit of observation	Firm-Year			Industry-Year		

This table reports the results on the relationship between ESG rating disagreement and ESG weight disagreement using tests at the firm-level and industry-level. All t-statistics are based on two-tailed tests and are calculated based on standard errors that are clustered by firm. All continuous variables have been winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. *, **, *** indicate statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.