Paying Attention to ESG:
Evidence from Big Data Analytics

Abstract

Environment, social and governance (ESG) issues are increasingly important to investors and ESG data have become the backbone of responsible investing. Yet, ESG ratings feature data quality issues with considerable dispersion among data providers. We propose a new measure based on *attention* to ESG issues using novel data of a firm's internet search intensity around ESG-related topics. We find large increases in a firm's attention to ESG-related topics predicts improvements in that firm's ESG ratings (Thomson Reuters' Refinitiv, MSCI's KLD, and RepRisk), Increased investor attention to ESG predicts changes in institutional investment positions, particularly sell-offs in stocks with poor ESG ratings. Studying the joint dynamics in attention to ESG among firms and their institutional investors offers some evidence that investors communicate with, possibly even influence, a firm's ESG actions.

Key words: ESG; Big data analytics; intent data; returns; institutional investors.

JEL Classification Codes: C55, G11, G23, G24, Q01,

Version : November 2020.

We are grateful to The Company for sharing their business-to-business intent data, especially an unnamed executive who answered many questions about the data science underlying the demographic and firmographic data at scale. The authors have secured access to this proprietary data under a non-disclosure agreement with The Company; there are no relevant and material financial relationships that bear on their research.

1. Introduction.

Businesses today are deeply intertwined with environmental, social, and governance (ESG) concerns. The "E" in ESG includes the energy a firm takes in and the waste it discharges, the resources it needs, and it encompasses carbon emissions and climate change; "S" is about the relationships a firm has and the reputation it fosters with people and institutions in the communities in which it does business, including labor relations and diversity and inclusion; and, "G" is the internal system of practices, controls and procedures a firm adopts to govern itself to comply with the law and meet needs of external stakeholders. In August 2019, the U.S. Business Roundtable strongly affirmed business's commitment to a broad range of stakeholders, including customers, employees, suppliers, communities as well as shareholders.[1] ESG-oriented investing has experienced a major rise with estimates now topping $30 trillion globally. According to a recent survey for the CFA Institute (Matos 2020), ESG implementation has been hampered to now because it has not been defined consistently, resulting in concerns about "greenwashing," or "rainbow washing," a false or exaggerated representation regarding how well aligned investments really are with sustainability goals. Matos further points to the fact that there are data quality issues linked to ESG ratings from commercial data providers and that these hamper inferences among scholars about the impact of ESG investing linked to real corporate change.

In this paper, we take a novel approach to understanding the consequences of ESG for corporate change by introducing new data on the attention paid to ESG by employees of the firms and their investors. The proprietary data come from a data analytics firm (which we will call "The Company") that specializes in measuring online/digital organization-level interest in business-related topics. The data cover a comprehensive subset of the U.S. firm universe and a wide range of digital content spanning thousands of topics across various themes of business, including but not limited to human resources, business strategy and operations, finance, marketing, enterprise and consumer technology, biotech,

[1] See "Statement on the Purpose of a Corporation" (August 2019) as signed by dozens of corporate CEOs around the world. See also Larry Fink's open letter, "A Fundamental Reshaping of Finance," BlackRock letter to CEOs (January 14, 2020) or Japan's Government Pension Investment Fund (GPIF), the world's largest pension fund, announcing revisions in 2017 to incorporate ESG issues, entitled "Fiduciary Duty in the 21st Century" (2019). The Global Sustainable Investment Review 2018 collates results from market studies of regional sustainable investment forums from Europe (Eurosif), U.S. (SIF), Japan (Sustainable Investment Forum), Canada (RIA Canada), Australia, and New Zealand (Responsible Investment Association Australiasia). The report is sponsored by Hermes Investment Management, RBC, and UBS.

engineering, construction and manufacturing. Their data are aggregated from thousands of partner media publishers, chief among which are household names, such as *Bloomberg* and *Wall Street Journal* in finance and *Laptop Mag* in technology. While many publishers are what one might consider news outlets, not all are. Some contribute technical content, policy-oriented white papers or video content. Each anonymized publisher contributes to the cooperative in return for access to the data provider's analytics products.

For each of the billions of web content interactions observed monthly across the publisher network, the firm categorizes the relevant topic, the location of the IP address and the organization associated with the IP address, where possible. The intended use of the dataset is to facilitate sales teams in finding sales leads among business customers – presumably, a prospective customer displaying elevated interest in the form of content interactions suggests the prospective customer might be likelier to buy a related product or service, shortening search costs and match rates. This data is part of advertising and marketing analytics called "intent data," and The Company is arguably the category leader.[2] We possess two versions of the data, which comprise a proprietary superset of the data made available on a special basis. The first is a weekly topic-firm interest index with scores dating back to May 2015 and extending through March 2019. This can be thought of similarly to Google Trends in that it is an index score between 0 (low reading intensity) to 100 (high reading intensity) produced at a weekly level, except that, unlike Google Trends, our data connects the reading level intensity *to a specific firm*. Second, we have a more granular, daily topic-firm-location counts of content interactions dating back to May 2016. The data currently occupy over 20 terabytes compressed.

The key innovation of our paper is to leverage The Company's intent data to construct a new measure of firm attention to ESG-related issues. We examine by hand more than 5,000 topics in our dataset and identify ESG-relevant topics. We place them into one of nine categories: Compliance, Corporate Governance, Customer Relations, Cybersecurity, Data and Sensitive Information Protection,

[2] A recent paper by Tong, Luo, and Xu (2020) includes a review of research across major marketing journals that study mobile marketing phenomena and consumer behavior changes using consumer hyper-context information (e.g. location, time, environment) to design personalized targeting ads. The earliest work includes that by Balasubramanian, Peterson, and Jarvenpaa (2002) and Barwise and Strong (2002).

Environment, Equality and Diversity, Labor Relations and Corporate Social Responsibility.[3] To the best of our abilities, these categorizations are chosen to mirror common classifications in industry ESG ratings.[4] This effort yields 323 topics, of which the largest categories are labor relations (63 topics), environment (46) topics) and corporate governance (29 topics). Examples of topics related to the environment include "Air Pollution," "Global Warming," "Climate Change," "Emissions."

Preliminary diagnostics suggest that our measures meaningfully capture a firm's attention related to ESG. In particular, we find that, in the category "Environment," firms in the Utilities and Mining, Quarrying, and Oil and Gas Extraction sectors spend the most time thinking about environmental issues. In the "Labor" category, the most time is spent by firms in the Educational Services and Health Care and Social Assistance sectors. This is consistent with the idea that firms in industries most exposed to these issues read more, a reasonable assumption for a measure of attention. Anecdotally, we study a firm's attention around a recent major ESG event; namely, Larry Fink released an open letter on January 14, 2020 to CEOs, which discusses the need of firms to focus improve their ESG performance. The letter focuses on environmental issues primarily, and to a lesser extent labor issues. Correspondingly, we plot the attention of firms with the highest percentage share by Blackrock as their investor versus firms with the lowest share or no ownership by Blackrock, finding that Blackrock-owned firms read more about the environment (and to a lesser extent labor) in the next two days, relative to firms with low or zero Blackrock ownership. For governance and data security issues, by contrast, there is no divergence between firms heavily owned by Blackrock and those not.

This quasi-exogenous experiment lends some plausibility to our working premise that a firm's internet search activity by employees can offer a meaningful measure of attention. But, in order to more rigorously validate these new indicators of ESG attention, we test the idea that ESG-related reading intensity is correlated with ESG-related actions in two settings. First, we examine firm-level ESG-relevant reading by *employees* and ESG-related outcomes. Second, we examine reading by *investors* and ESG-related outcomes. Finally, we examine the co-movement between investor reading and firms'

[3] This category covers topics almost exclusively on social issues, other than those focused on labor, customer and diversity. It includes sub-topics like corporate philanthropy and community engagement.

[4] See, among others, the report by Tim Koller, Robin Nuttall, and Witold Henisz, "Five Ways that ESG Creates Value," *McKinsey Quarterly* (November 14, 2019). Also, Table 1 in Matos (2020) offers a useful classification.

reading. As much as firms are involved in acting in a socially responsible way, it is often argued that investors prefer firms that are more socially responsible and even influence firms directly on such matters (Hoepner, Oikonomou, Sautner, Starks, and Zhou, 2019; Dimson, Karakas, and Li, 2015, 2019).

There are many ESG-related metrics available to financial investors employing a variety of approaches (Matos, 2020). For data availability, we choose three of the most popular ones: *Refinitiv* (previously, "Asset4"), a Thomson Reuters database; *RepRisk*, built by a Swiss company that leverages artificial intelligence (AI) and machine learning for ESG risk measurement; and, Morgan Stanley Capital International (MSCI) ESG Research's *KLD* (formerly, "KLD Research & Analytics"). All three are used in academic studies and notably differ in their implementation methods. Whereas KLD and Refinitiv are relatively low-frequency measures employing teams of analysts to write detailed reports and perform company specific analysis, RepRisk, as mentioned, is premised off of a machine-learning approach. Its natural language processing technology identifies controversial news and related firms.

There are three possible relationships we identify between firm attention and ESG ratings. First, rank-and-file employees of an organization may consume information about ESG merely as a hobby or a sideshow. In this way, the time-varying intensity of attention allocation toward ESG issues may have no bearing on the ESG-related or overall performance of the firm. This is our working null hypothesis, which we could reject in one of two ways. First, we may find a negative relationship between ESG reading intensity and ESG performance. This could arise if organizations read about ESG ahead of impending negative environmental or social news. Second, we may uncover a positive relationship between ESG reading and ESG performance if organization members read about ESG either ahead of publicly-released positive news, or if they consume information that they might need (legal advice, investor relations, or just topical knowledge) ahead of actions they specifically take to improve the firm's ESG reputation. Of course, we cannot distinguish between these two, as it is hard to verify that the consumption of information leads necessarily to any specific action.

Our first tests examine ESG reading intensity and Refinitiv ratings. In a firm-year panel, we test whether firms which tend to have high reading activity tend to see *improvements* in ESG score. Across a variety of specifications, we find that higher reading activity increases the ESG combined score as produced by Refinitiv. Importantly, this holds for specifications saturated by year fixed effects, industry fixed effects and firm fixed effects, and controlling for high reading intensity levels that are unrelated to ESG. The economic magnitude is large: A one standard deviation increase in ESG reading is associated with an average increase in the ESG combined score of 9.16%, within firm and within time. Furthermore, we find reading intensity that is unrelated to ESG tends to either be not at all or negatively correlated with a firm's future ESG performance. This suggests our topic selection is meaningful. Refinitiv also has two sub-components of ESG combined score: the overall "ESG" score, in which Refinitiv scores the firm based on perceived ESG practices, and a "Controversies" score, which refers to the perception of the firm in global media sources. The ESG Controversy Prediction from Refinitiv uses a combination of supervised machine learning and natural language processing to train an algorithm. The algorithm classifies whether articles contain reference to 20 ESG controversy topics and yields a probabilities score for each of them. We find that the positive effect of ESG attention to ESG scores arises primarily from the "Controversies" score.

RepRisk, like the Refinitiv Controversies score, is based on global media sources and identifies negative news surrounding a firm. The benefit of RepRisk is that it is systematic and is updated on a monthly basis. Here, we seek to predict the "RepRisk Index" score, which measures the severity of ESG-related risk in the current month. At the firm-month level, we find that firms with large increases in ESG-related reading tend to be associated with lower RepRisk scores, an indication that the controversies surrounding a firm tend to decrease when a firm pays attention to ESG-related issues. A one standard deviation increase in ESG reading is associated with an average decrease in the ESG combined score of 5.14%, within firm and within time.

Our final firm-level test examines ESG reading intensity and KLD ratings. We first use the adjusted KLD score widely used in the literature (Deng, Kang, and Low, 2013; Servaes and Tamayo, 2013, for example). At the firm-year panel, we show there is a positive cross-sectional relationship between firm's ESG reading intensity and the adjusted KLD score, but there is no relationship when conducting analysis within-firm. We discuss this finding in the light of concerns raised by Chatterji, Levine, and Toffel (2009), who notes that KLD's ratings are not optimally using publicly available data. Another possibility, however, is that the naïve average metric adopted in the literature is sub-optimal.

Thus, we present two alternative formulations of the KLD rating which statistically seem more closely linked to firm attention.

In our institutional investor-related tests, we build measures of reading intensity about ESG topics among employees of major asset management firms that parallel those for the industrial and services firms above. Here, we examine whether investors that read more about ESG issues are likelier to invest in firms with lower ESG-related risks. We find evidence to support this conjecture. In our portfolio-level results, we show that when investors read more about ESG issues, they tend to hold portfolios of stocks that have better ESG ratings or with lower ESG risks. When we turn to a more granular investor-stock-quarter analysis, we uncover that, when institutional investors read more intensively about ESG issues, they are more likely to invest in or are less likely to sell (especially sell-off completely) stocks that have better ESG ratings. The result holds for specifications saturated by investor × quarter fixed effects, a variety of stock characteristics and controlling for investor's reading intensity levels that are unrelated to ESG.

In our final tests, we study the relationship between investor's and firm's ESG-related reading intensity, what we call their joint dynamics. To that end, we calculate a cosine similarity measure between pairwise ESG reading intensity across ESG topics. Investors are more likely to invest in stocks that have similar ESG reading intensity by topic, especially when investors read intensively about ESG issues. The result holds for specifications saturated by fixed effects - investor × quarter, stock × quarter, investor × stock - and controlling for investor's reading intensity levels that are unrelated to ESG or similarity for non-ESG issues. The association between investor's and firm's ESG reading at a topic-month level shows that when top five investors experience an increase in reading intensity on a specific ESG topic, there is a 2.9% higher likelihood that the firm's reading also jumps on that topic. Other investors outside the top five are only associated with a 1.4% higher likelihood of an increase in reading.

These findings contribute importantly to the existing literature on ESG investing, which expressed many concerns over the measurement of ESG. In an important review, Matos (2020) documents two trends: (1) the rise of interest in ESG issues, particularly those related to climate change, and (2) a general lack of consensus among academics as to what the core issues of ESG are and how it should be measured. Following this latter thrust, Gibson, Krueger, Riand, and Schmidt (2019) find that the average correlation between overall ESG ratings of six different ESG data providers was less than 50%. Chatterji, Durand, Levine, and Touboul (2016) attribute the observed disagreement in ESG ratings to a lack of shared view of what it means for a firm to be socially responsible. Indeed, Gibson, Krueger, Riand, and Schmidt (2019) show how legal origin of ESG rating firms matter: civil-law-based ESG data providers have strong views on labor/social issues, while common-law-based providers focus on shareholder right protections. Berg, Kolbel, and Rigobon (2020) point to scope, measurement, and weights as sources of divergence among ESG ratings.

Our novel attention-based approach, currently not implemented by industry practitioners to the best of our knowledge, provides us a unique window into the disagreement among ESG ratings data. Our approach allows us to determine which ratings correspond most strongly to a firm's *attention* to ESG issues. This allows us to discern a good ESG rating with the working assumption that a good ESG rating should correspond to active efforts made by the firm's employees to research the matter. In terms of investor attention to ESG, that our measure of attention seems to explain subsequent investor actions on stocks with certain ratings can be viewed as a revealed preference measure of how – if at all – these ratings are being used. Recently, Cao, Titman, Zhan, and Zhang (2020) argue that investors pay attention to ESG as implied by their holdings of ESG-linked stocks. Though different in approach, our paper and theirs are closest in purpose; we measure ESG attention directly and, in doing so, validate their approach, while pushing the empirical envelope through new data.

2. Data.

2.1 Intent data.

For this study, we obtained proprietary data on internet research activity from a business-to-business (B2B) "intent" data provider, The Company. Intent data refers to a recent development in data analytics aiming to gauge a prospective business customer's buying interest based on patterns of web content consumption. The premise of intent data is that if an economic agent consumes more web content related to a particular topic, this agent might have elevated interest or "intent" in procuring a related product or service. The company who supplied their data tracks organizational level (tracked by web domain, such as microsoft.com) interest in specific topics at specific locations (for example, Microsoft's interest in "Python" in Redmond, WA).

The Company whose data we use is a leading provider of intent data. At the most basic level, a single company may observe its audience members and their behaviors on its website. This proprietary dataset leverages not only visits to a firm's own website, but rather orchestrates a cooperative of contributors under a "give-to-get" model. Co-op members consist of thousands of mainstream business media sites such as *Wall Street Journal*, *Bloomberg*, *Forbes*, *Business Insider* and *CBSi*, along with more specialized groups of sites such as *1105Media*, *ITCentral Station* and *Questex*. Most sites are anonymous but span a wide range of business functions, such as technology, finance, marketing, legal, human resources, engineering and manufacturing and general business. Co-op members participate because they receive some of the services and especially the data analytics The Company sells.

Co-op members contribute to the pooled dataset via a technology mechanism which shares information about web content consumption, including the external IP address of the network originating the HTTP request and the URL of content accessed. This data is then filtered into domain, location and topic. To do so, The Company performs two major steps. First, it assigns the IP address to a work email and web-domain using an ensemble of methods. Additionally, a user profile is generated, via consent-based and anonymized third-party cookies, which, in combination with the external IP address, allows the data provider to associate a domain with the profile, when such an association can be inferred. This association is likely to convey where someone works or a work-device. For example, if users from an IP address consistently log onto a publisher website from the same work email domain, this gives a strong indication this is a business address belonging to that company.

The content is tagged for topics. The Company operates a supervised learning algorithm using a hand-picked set of training manuals which have been labeled for topics the company aims to study. For example, to chart interest in "Cloud Computing," they have assembled a set of 80 to 100 articles that has been labeled as being pertinent to cloud computing. Topics are chosen by The Company either as a result of publisher and customer requests or according to The Company's view on business-relevant topics and issues.

Table 1 presents summary statistics on our intent dataset coverage. Over the course of the five years in our period of analysis from 2015 through mid-2019, the number of topics identified by The Company has nearly tripled from 2,462 to 6,765. Panel A exhibits a similar pace of growth arises in the number of web-domains – or business "addresses" - that The Company tracks from 1.7 million in 2015 to 6.9 million by the end of our sample. To get a proper sense of the data, we also report the number of domain-mapped business-related interactions per day which reaches a peak of 686 million in 2017 across the 4.3 million domains and for 3,589 different topics as of that year. Panel B lists the topic taxonomy as of 2019 by themes as defined by The Company. This range of topics indicates the breadth of internet research interests across the domains The Company tracks.

Most of our analysis uses what The Company calls their "Spike" score. [This is our name for The Company's index score and not that of The Company itself.] A firm's Spike score is a weekly index aggregate which measures a firm's topic-level interest and it dates back to May 2015. The score runs from 0 to 100, with a score of 50 representing no increase in interest, a score above 50 representing a mild increase in interest, and a score of 60 representing a significant increase in interest. Conversely, a score below 50 represents a mild decrease in interest. Scores are produced for those topic-domains in which there is a threshold number of observations in the first and last 3 weeks of a 12-week window. A high Spike score represents a large increase in the prior 3 weeks relative to the preceding twelve weeks, accounting for other firms' relative increases in this same topic over this same period. This last step is important because aggregate interest in a topic might have increased due to mechanical changes in the topic taxonomy or due to an increase in the supply of publisher content without an actual increase in unique, firm-specific interest. This, and a variety of other proprietary adjustments, facilitates the detection of genuine bursts in reading interest, rather than just mechanical increases.

2.2 Defining ESG-related topics.

From the several thousands of topics provided, we hand pick topics most relevant to ESG. We pick 323 ESG topics in total and classify them into nine categories: Compliance, Corporate Governance, Customer Relations, Cybersecurity, Data and Sensitive Information Protection, Environment, Equality and Diversity, Labor Relations and Corporate Social Responsibility. Our nine categories are in the spirit of the common classifications in industry ESG ratings. But we recognize that there is some discretion in our choices. In our main analysis, we decompose topics into four ESG categories: Environment, Labor (including Labor Relations, Equality and Diversity), Social (including Customer Relations, Corporate Social Responsibility) and Governance (including Compliance, Corporate Governance, Cybersecurity, Data and Sensitive Information Protection). Appendix A outlines how we assign ESG topics to the intent data.

In Panel A of Table 2, we show the number of topics within each category we classify. The category "Labor Relations" contains the highest number of topics while "Compliance" contains the lowest number of topics. The relative differences between compliance and labor relations could be ascribed to one of three possibilities: (1) the composition of The Company's cooperative (the publishers who contribute content), (2) the composition of content on the internet, (3) or the composition of the firm's topic engine, which might have simply more topics tracked in these areas. In Panel B of Table 2, we show ten examples of topics from the proprietary dataset within each of the four ESG dimensions to illustrate the categorization approach. "Labor" includes topics of both Labor Relations and Equality and Diversity. We list some of the top topics in each category. For labor, the top 5 are "Diversity Recruiting," "Employee Safety," "Equal Employment Opportunity," "Equal Pay/ Comparable Worth," and "Gender Equality."

At first glance, three of these topics seem somewhat related; diversity, gender equality, and equal employment opportunity all refer to equity regardless of race or gender. However, The Company implicitly accounts for correlations between topics through its topic engine. That is, if an article pertains to both diversity recruiting and equal employment opportunity (EEO) equally, the article will be given a weight of 50% for both topics. In this way, double-counting concerns are mitigated. We report the number of topics, domains, and domain-mapped business-related interactions per day among ESG topics in Panel C of Table 2, the counts for which can be compared against those of the universe of all topics reported in Panel A of Table 1. ESG topics as we have defined them represent a substantial fraction of all interactions per day ranging from a low of 6.74% in 2016 to a high of 9.40% in 2017.

In Panel D of Table 2, we list the ten industries which have the highest percentage of reading intensity across each of the four ESG dimensions during our sample period. We define the percentage of reading as the total record of topics in that dimension divided by total record of all topics. We define industry as first two digits of NAICS code. The first dimension we show is "Environment." The industry that reads with the most intensity about ESG matters is Utilities, followed closely by Mining, Quarrying, and Oil and Gas Extraction. These industries are those which pay much attention to air and water pollution, climate change and sustainable energy. The second dimension we show is "Labor," which includes topics of both Labor Relations and Diversity and Inclusion. The industry that reads most intensely on these topics is Educational Services, followed by Health Care and Social Assistance, and Accommodation and Food Services, industries which are all labor-oriented. The third dimension is "Compliance," of which the Finance and Insurance and Professional, Scientific, and Technical Services industry reads with the most intensity. The last dimension we show is "Data and Sensitive Information Protection." The top industries are Finance and Insurance, Professional, Scientific, and Technical Services, and Health Care and Social Assistance which are all sensitive to data and information privacy issues. The industry reading results show that our ESG measures can reliably rank industries in terms of their attention to ESG.

2.3 ESG ratings and other data.

Our first source of ESG rating data is *Refinitiv* (formerly, Thomson Reuters' Asset4 ESG database). Refinitiv collects ESG-related information of publicly traded firms from public sources, such as annual reports, Corporate Social Responsibility (CSR) reports and non-governmental organization (NGO) websites. Then Refinitiv captures and calculates over 450 company-level ESG metrics and combines them into ten main categories. The weighted average of ten category scores finally formulates ESG score which reflects firm's annual relative ESG performance. Refinitiv also provides ESG combined score which is discounted for significant ESG controversies. In this paper we rely on ESG combined, ESG, and ESG Controversies scores to measure firm's ESG performance. Higher scores indicate better ESG performance. And we use data from May 2015 through year-end of 2018 for our analysis. We start from 2015 as this is the first year covered by our un-named intent data firm.

Then we obtain monthly ESG-related risk data from *RepRisk*. RepRisk differs from Refinitiv and KLD (discussed below) in that it relies more on a computerized, systematic approach. RepRisk scours the internet for regulatory filings and news articles in multiple languages, scouring tens of thousands of sources. When its algorithms screen an event damaging the firm's ESG reputation, it applies a human analyst to verify the information and enter it into its database. The data are then used to compute a monthly RepRisk index per firm. We mainly use monthly current RepRisk Index (what we will call *Current RRI*) to measure firm's ESG-related risk and a higher index means more exposure to ESG-related risk. The sample period starts from May 2015, which is the first month covered by the Company. It ends in August 2018, well before the end of our intent data series end, but the last month we can obtain data from RepRisk as of writing.

Finally, we obtain firm-level rating data by MSCI ESG's *KLD*, which mainly rates firms based on a wide range of strengths and concerns across seven categories: community, diversity, employee relations, environment, governance, human rights, and product. Specifically, each KLD category has a variety of areas, and if a firm has strengths/concerns it will be given one point in that area. We use data from 2015 to 2018, the last year covered by Wharton Research Data Services (WRDS) as of writing.

Our institutional holdings data comes from FactSet Ownership because we can identify both investor's and firm's web domains and can link to The Company's proprietary dataset. We obtain company identifiers and financial data from Standard & Poor's Compustat and stock market data from the Center for Research on Security Prices (CRSP).

3. Assessing the ESG reading intensity of firms.

Having introduced the Spike Score measures among ESG topics across firms and years and having outlined their summary statistics, our first step is to provide evidence that these measures meaningfully capture a firm's attention related to ESG. We specifically examine firm-level ESG-relevant reading by a firm's employees and ESG-related outcomes, as defined by our Refinitiv, RepRisk, and KLD ESG index scores.

3.1. Distribution of the Spike Score.

In Table 3, we report the distribution of weekly Spike Score for all CRSP firms by year. In each year, we report a variety of quantiles of the distribution - 25^{th}, 50^{th}, 75^{th}, 90^{th}, 95^{th}, 99^{th} - of Spike scores. Panel A shows distributions for all topics and, in Panel B, we show them for ESG topics only. Recall that The Company regards a score above 60 as a significant interest in a topic-domain, as measured based on the prior three weeks relative to the 12 weeks before that with some adjustments. From the distributions, we can see that a Spike Score in higher ends of distributions (95^{th} or 99^{th} percentiles) are stable across different years and that they are similar for all topics or ESG topics. A Spike Score of 80 as a threshold for high interest in a topic in a given firm-week lies between the 95^{th} and 99^{th} of the firm-week distribution. While admittedly an arbitrary threshold, we choose the number of Spike scores that are at least 80 to capture high reading intensity or attention. Hereafter, Spike refers to Spike Score 80.

3.2 ESG reading intensity by firm and ESG rating by Refinitiv.

We first obtain ESG ratings data of all covered firms from Refinitiv, with ISIN the identifier of each company. Then, for each ISIN we get GVKEY and websites from Compustat. Finally, we merge with the intent data using the firm's domain.[5] We require a non-missing ESG combined score and a firm's fiscal year-end at December to better align with the intent data firm's annual data. Our match delivers 7,582 firm-years among 2,164 unique firms.

Panel A of Table 4 shows summary statistics of three original ESG scores by Refinitiv and the main variables used in this regression analysis. The combined scores for Refinitiv averages at 45.265 on the 0 to 100 scale with a standard deviation of 16.238. There is even greater variation across firm-years for the sub-component core ESG score as well as the Controversies score. In Panel B of Table 4, we test whether and how a firm's ESG-related reading intensity is associated with a firm's ESG rating by Refinitiv. In this panel, ESG-related reading intensity is captured by number of Spike scores that reach at least 80 in a given firm-year and we take logarithm of that count as the main variable, or $Log(1+Spikes^{ESG})$. The regression specifications add year fixed effects across all columns, industry fixed effects in one and firm fixed effects across most others.

The summary statistics for $Log(1+Spikes^{ESG})$ imply a mean of 3.912 and associated interquartile range of 3.466 to 4.804. This mean represents the equivalent of 48.9 Spike Scores of 80 among the 2,164 firms in a given year with an interquartile range from 32 to 121.5. For first four columns of Panel B, we use the level of ESG combined score ($Refinitiv^{Combined}$) as the outcome variable of interest. Recall that the combined score is the firm's relative ESG performance across more than 50 metrics along with its ESG controversies overlay. In Model (1), we add industry fixed effects to control any industry

[5] Domain is the clean version of website. For example, we clean "www.google.com" or "http://www.google.com" to "google.com". Domain is more unified and accurate to use for merging with firms.

invariants. The estimated coefficient of *Log(1+Spikes*^{ESG}*)* is positive and statistically significant at the 1% level. When adding firm fixed effects in Model (2), the estimated coefficient of *Log(1+Spikes*^{ESG}*)* remains positive, abates somewhat in magnitude, but it is statistically significant at 5%. In Model (3), we further control non-ESG reading intensity *Log(1+Spikes*$^{Not\ ESG}$*)* and the estimated coefficient of *Log(1+Spikes*^{ESG}*)* becomes larger and more reliable in terms of statistical significance. The economic magnitude implied by this estimated coefficient of 1.108 is also large: a one standard deviation increase in ESG reading is associated with an increase in the ESG combined score of 1.484 (1.339 × 1.108), which represents 9.16% of its standard deviation (16.238). In Model (4), we decompose ESG reading intensity by Environment, Labor, Social and Governance to understand what drives the positive association. The result shows that, for the ESG combined score, the Social category dominates.

In Models (5) and (6), we study the level of the ESG core (*Refinitiv*^{ESG}) and ESG Controversies scores (*Refinitiv*^{Contro}) as dependent variables, respectively. The coefficient of *Log(1+Spikes*^{ESG}*)* remains positive but it is only statistically significant and at the 5% level for ESG controversies score. The above results imply that higher ESG-related reading intensity is associated with improvements in the ESG combined score, and that, in turn, it is concentrated in the association with the in ESG Controversies score. Moreover, the component of the Spike linked to the Social category is that most prominently associated with the ESG combined score.

3.3 ESG reading intensity and RepRisk risk ratings.

We next obtain monthly RepRisk index (*Current RRI*) score for all covered firms with the identifier RepRisk ID and then merge these data with monthly The Company data using the following methods. First, we obtain the ISIN for each RepRisk ID from RepRisk, where possible. Then, we merge with The Company data using the same way as in Table 4. Finally, we have 59,413 firm-month and 1,735 unique US firms.

In Panel A of Table 5, we show summary statistics of *Current RRI* and main variables used in the analysis. The mean *Current RRI* is 12.827 with a standard deviation of 12.097. The distribution of scores across firm-months is left-skewed with lots of zero values (at least 25% of the observations) and a maximum value of 55. In Panel B of Table 5, we test whether and how a firm's ESG-related reading intensity is associated with a firm's ESG-related risk. We use *Current RRI* as the outcome variable and add month fixed effects across the first four columns. Models (5) and (6) are firm-year specifications with year fixed effects and are directly comparable with those in Table 4. In Model (1), we only add industry fixed effects and the coefficient of *Log(1+Spikes^{ESG})* is positive and statistically significant at the 1% level. Our explanation is that RepRisk is a firm-specific index for ESG-related risks and firms with relatively high ESG risks may pay more attention and efforts to ESG, which drives the positive cross-sectional association. In Model (2), we add firm fixed effects instead and the coefficient of *Log(1+Spikes^{ESG})* is negative and statistically significant at the 1% level. This is consistent with the positive sign in Model (6) for the Refinitiv Controversies score. The economic magnitude implied is notable: A one standard deviation increase in ESG reading is associated with an average decrease in *Current RRI* of 0.337 (1.037 × 0.325), which represents 2.78% of its unconditional standard deviation. In Model (3), we control non-ESG reading intensity and the result is similar. In Model (4), we decompose ESG reading intensity by Environment, Labor, Social and Governance to understand what drives the negative association. The result shows that "E" and "S" categories matter the most.

To make a head-to-head comparison with annual Refinitiv or KLD rating, in Models (5) and (6) we aggregate the firm-month observations to firm-year by using the year-end *Current RRI*. The result is similar except that when controlling non-ESG reading, the coefficient of *Log(1+Spikes^{ESG})* becomes insignificant. This is somewhat surprising given the resilience of the measure in Table 4. It does imply that there is important information in the higher-frequency monthly RepRisk scores that is lost by constructing a lower-frequency annual level of granularity.

Overall, the above results show that ESG-related reading is negatively associated with the level of RepRisk Index (*Current RRI*), which means ESG-related reading can mitigate firm's ESG-related risk. The mitigation mainly comes from Environmental and Social categories of reading intensity.

3.5 ESG reading intensity by firm and ESG rating by KLD.

We use KLD (also known as MSCI *ESG KLD STATS*) as our third source of ESG rating data. We obtain all covered firms in KLD and get company identifiers from CRSP/Compustat Merged Database. We match with The Company domain identifiers using the same way as we did in Table 4. For these tests, we have 2,462 unique firms.

Recall that KLD reports a wide range of strengths and concerns across seven categories: community, diversity, employee relations, environment, governance, human rights, and product. Each KLD category has a variety of sub-areas, and if a firm has strengths/concerns it will be given one point in that area. We sum all strengths and concerns for any firm-year to obtain a count of total strengths (*Str*) and total concerns (*Con*). Because the maximum possible number of individual strengths and concerns in each category may change over time, we follow previous studies, such as Deng, Kang, and Low (2013) and Servaes and Tamayo (2013), to construct the adjusted KLD score *KLD1* which equals *(Str-Con)/(n_Str+n_Con).* The variables *n_Str* and *n_Con* are the maximum possible number of strengths and concerns across categories and sub-areas, respectively. We also construct another version of adjusted KLD score, which is defined as *(Str-Con)/(Str+Con).* We think of this as an analogous measure of relative strength as *KLD1* but captured on the "intensive margin," in which we exclude those counts of strengths and concerns for which there are zero values. We use two different methods to calculate this relative strength at the intensive margin when there are no strength or concern for any firm year: the KLD score can be zero (*KLD2*) or missing (*KLD3*), which allows for many fewer firm-year observations as a result.

In Panel A of Table 6, we show summary statistics of three versions of the KLD score and main variables used in analysis. As expected, the mean relative strength measures for *KLD2* or *KLD3* are much higher at 0.558 and 0.562, respectively, than those for *KLD1* at 0.024. The standard deviations are similarly much higher across firm years In Panel B of Table 6, we test whether and how a firm's ESG-related reading intensity is associated with adjusted KLD scores. We add year fixed effects across all columns. For first three columns, the dependent variable is *KLD1*. In Model (1), when we add industry fixed effects, the estimated coefficient of $Log(1+Spikes^{ESG})$ is positive and statistically significant at 1%. When adding firm fixed effects instead in Models (2) and (3), the coefficient of $Log(1+Spikes^{ESG})$ becomes insignificantly negative. In Models (4) and (5), we repeat the analysis with *KLD2* and *KLD3*. The coefficient of $Log(1+Spikes^{ESG})$ is positive but only statistically significant at 1% when using *KLD3* as the outcome variable. The economic magnitude here implied is large: A one standard deviation increase in ESG reading is associated with an average increase in *KLD3* of 5.990 (1.116 × 5.367), which represents 15.85% of standard deviation of regression residuals (37.780). In

Model (6), we decompose ESG reading as in Tables 4 and 5, where we see the Governance category dominates the overall effect.

The above results suggest that KLD score may not capture a firm's attention or efforts in ESG well, especially when it comes to within-firm analysis. It could be due to the fact that KLD rates firms based on a variety of indicator variables which provide less information compared to numeric rating. For example, in a given strength area, if one firm puts more efforts and resources to improve its profile than previous years but still fails to reach the standard or threshold set by KLD, it will be rated as zero. The same applies to KLD concern areas. Chatterji, Levine, and Toffel (2009) also raises similar concerns and argues that KLD's ratings are not optimally using publicly available data.

We run a battery of specification checks which we report in an appendix for brevity.[6] First, we ran these analyses not as *level regressions* but rather as change specifications. This might actually be more intuitive than running a level specification in that the Spike score is a measure of a firm's attention relative to the recent past. That is, one might interpret it as a change or innovation measure as well. We find that when viewed as a change measure, we get roughly the same inference as before.

We also conduct analysis using what The Company calls its Daily Aggregates file, which is the input file that underlies the weekly 'Spike' score. In this file, we can count the exact number of records read by the organization pertaining to a particular topic, giving us a count measure rather than a 'spike' measure. This count measure is possibly easier to interpret. However, we do not use this as our primary measure for two reasons. First, it has a short sample period, going back to mid-2016. Therefore, we lose a crucial year our relatively short sample period and in the assessment of our annual metrics. Second, there are arbitrary flaws pertaining to changes in the topic interest model or publishers which contribute to the underlying dataset. Thus, it is not obvious that a raw record count can be comparable from one period to the next.

That said, using this alternative version of the data, we find analogous result that roughly confirm our main analysis. We present two specifications: the percentage of reading related to ESG, and secondly deflating the record count by assets. The former measure, the relative allocation of

[6] See tables in Appendix B.1 and Appendix B.2.

attention to ESG, is positively related to the Refinitiv score, while the latter measure seems more strongly related to KLD. One possible interpretation is that RRI and KLD are measures that are biased with respect to the firm-size distribution, whereas Refinitiv is less so.

4. Assessing the ESG reading intensity of investors.

In this section, we assess whether greater ESG reading intensity by employees of asset management holding firms is linked to quarterly changes in the investment positions of institutional investors with respect to stocks with high ESG risks and/or with better ESG ratings. Our approach follows in two stages. The first stage evaluates links between ESG reading and averages of ESG scores among the stocks held by the asset management firm at the investor-quarter level. We call this portfolio-level ESG performance. During quarters in which there are jumps in ESG reading intensity, do we observe changes in the percentages of stocks held by the firm with high ESG scores. The second stage pursue a more granular analysis at the investor-stock-quarter level with over 5 million observations in which we examine whether an asset management firm increases/decreases its position in a stock in a way that is associated with a stock's ESG score *and* in a way interacted with the intensity of the firm's ESG reading intensity. We refer to this changes of positions at the stock-quarter level as the asset management firm's trading decisions linked to ESG.

4.1 Investor's ESG reading and portfolio-level ESG performance.

We first obtain holdings data from FactSet which provides the websites for the overall holding firms. After we get investor-level domains, we merge holdings data with The Company to get an investor's quarterly ESG reading intensity. We use all common stocks in CRSP for this analysis. Our final sample starts from the third quarter of 2015 (first complete quarter covered by The Company) and ends at the fourth quarter of 2018 (last quarter of FactSet). The ESG ratings of stocks are calculated before the start of the quarter in which the investors conduct ESG reading.

Based on this sample, we first assess the relationship between investor's ESG reading intensity $Log(1+Spikes^{ESG})$ and ESG performance of an investor's portfolio. We take an equal-weighted approach to calculate portfolio-level ESG ratings because we do not want the results to be mainly driven by large-cap stocks. We present the results in Table 7. Across all columns, the main explanatory variable is investor's ESG reading intensity *Log(1+Spikes^{ESG})*. In each case, we control for that investor's non-ESG reading intensity, as well as quarter and investor fixed effects. In Models (1) to (5), we use equal-weighted ESG ratings following those that were featured in Table 4, 5 and 6. Model (1) indicates that investor's ESG reading intensity is positively associated with adjusted KLD score (the version commonly used in literature). Models (2) to (4) feature the ESG Combined Score and ESG Score by Refinitiv. The coefficients of *Log(1+Spikes^{ESG})* are statistically significant at 1% for adjusted KLD score and Refinitiv ESG Score, and at the 5% level for Refinitiv ESG Combined Score. The magnitude of the effect is noteworthy: based on Model (2), a one-standard-deviation increase in reading intensity is associated with a 1.698 (1.268 × 1.339) higher Refinitiv Combined score, or 10.5% of the standard deviation across investment firm-quarters. Surprisingly, we do not find statistically significant results for Refinitiv ESG Controversies Score as we did in Table 4 at the firm-year level of analysis.

Model (5) also reveals that the large increase in ESG reading by the asset manager is not associated with the *Current RRI* from RepRisk. In Model (6), however, we use an equal-weighted *Peak RRI*, which is the maximum of *Current RRI* in the last two years. One can think of this as a high-water mark on "reputation" in ESG risks that can carry forward over time. The coefficient on *Log(1+Spikes^{ESG})* is negative, but insignificant. In Model (7), we transform the outcome variable to be a percentage of stocks that have positive *Peak RRI* among all those stocks held by the asset management firm that quarter. This threshold measure is negatively associated with *Log(1+Spikes^{ESG})* in a way that is consistent with our findings in Table 5 – asset management firm employees increase ESG reading intensity during quarters in which they have a relatively high fraction of stock holdings with high ESG risks. In Model (8), the outcome variable is transformed to be the percentage of stocks that have *Peak RRI* that is larger than or equal to 50. We use the threshold 50 because it is used by RepRisk to classify stocks with high ESG risks. This is the equivalent to the measure in Model (7) but with an additional threshold. The results indicate that the coefficient of *Log(1+Spikes^{ESG})* remains negative, but it is only statistically significant at 1% in Model (7).

4.2 *Investor's ESG reading and trading decisions.*

To give more direct evidence than our portfolio-level tests, we next turn to investor-stock-quarter levels of analysis using the same sample as above. We define several variables for this analysis.

A quarterly measure of investment, $Invest_{ijt}$, is defined as log change in the dollar value of investor *i* holdings for stock *j* in quarter *t* and can be computed as:

$$Invest_{ijt} = log(1 + Holdings_{ijt}) - log(\ 1 + Holdings_{ijt-1}(1 + Return_{jt})),$$

where the term $Holdings_{ijt}$ is dollar holdings by investor *i* for stock *j* at end of quarter *t*. A firm's stock return is represented by $Return_{jt}$.This captures changes in positions of existing holdings, but we also define different types of investment. "Selloff" is defined as a liquidation of all shares in an existing position. "Decreases" are defined as more than one percent decreases in dollar holdings of a stock. A "Hold" is defined as any change in dollar value of a position within one percent change of dollar holdings as of the beginning of the quarter. "Increases" is defined as more than one percent increase of dollar holdings. Finally, a "Pickup" is defined as a *de novo* investment in a stock that was not held by the investor last quarter. Each of these are defined as indicator variables equal to one if the change in position is equivalent to the definition above and zero, otherwise.

In Table 8, we examine how investor's ESG reading and stock's ESG rating affect investor's overall investment as well as different types of investment. We add various stock characteristics by last quarter-end that could influence an investor's trading decisions: the trailing quarterly stock return and volatility, its market capitalization, momentum, gross profitability, and book-to-market ratio. Volatility is computed from daily returns. Momentum is the return of the over the past year, skipping one month. Gross profitability is the net income over assets, while book-to-market deflates book value by market cap. After requiring stock-level control variables to be non-missing and the stocks to be held by the investor last quarter, there are 6,359,854 investor-stock-quarter observations. The sample is diminished when merging with ESG ratings data, but allows for 3,831 different institutional investors and 2,831 unique stocks.

Panel A exhibits summary statistics of main variables in the analysis. Across the nearly 6.4 million stock-investor-quarter observations, the average dollar value of a position change of the typical investor in a typical stock is negative at -1.574 percent per quarter during 2015-2019. There is considerable variation with a standard deviation of 4.286 percent per quarter. We standardize all acronyms for ESG ratings for this analysis and name them $Score^{ESG}$ in Panel B – the heading in each of the four columns represent the ESG rating score that applies. We run regressions for the overall investment and, in these tests, the stocks are only those held by the investors last quarter. The variable of interest is $Score^{ESG} \times Log(1+Spikes^{ESG})$. Across all columns we control non-ESG reading intensity by investors, stock characteristics and investor-quarter fixed effects. In Model (1) we present result for adjusted KLD score (*KLD1*). The coefficient of $Score^{ESG} \times Log(1+Spikes^{ESG})$ is positive at 6.104 and statistically significant at 1%. The result is similar when we use ESG Combined Score by Refinitiv in Model (2). When we use *Current RRI* or change of *Current RRI* in Models (3) and (4), the coefficient of $Score^{ESG} \times Log(1+Spikes^{ESG})$ becomes insignificant, although there is some evidence that investors try to reduce investments in stocks with high ESG risks, but most especially those with increases in ESG risks – Model (4) has a coefficient of -3.368 of $Score^{ESG}$ for *ΔRRI* with a robust *t*-statistic of 4.26. The weak link with the Spike score is perhaps indicative that by some measure of revealed preference, investors do not use RepRisk scores as they do KLD or Refinitiv.

In Panel C, we use different types of investment as outcome variables to understand what drives the result. In these regression tests, we use ESG^{Zscore} which is the sum of the standardized adjusted KLD score and ESG Combined Score.[7] Models (1) to (4) presents results for stocks held by the investors last quarter. Out of 4,796,856 observations, there are 528,141 "Selloff" events, 2,287,602 "Decreases," 1,071,621 "Hold" events, and 1,437,633 "Increases." The results suggest that "Selloffs" dominate. The coefficient of $Score^{ESG} \times Log(1+Spikes^{ESG})$ is negative and statistically significant at 5%, indicating that when stocks have better ESG performance *and* investors have strong ESG interest (proxied by high ESG reading intensity), investors are less likely to exit completely from those positions.

In Model (5), we present result for "Pickups" of stocks that were not held by investors last quarter. Among 39,241,287 qualifying observations, there are 469,042 "Pickup" events. The coefficient of $Score^{ESG} \times Log(1+Spikes^{ESG})$ is positive and statistically significant at the 1% level, suggesting that when stocks have better ESG performance and investors have strong ESG interest, investors are more likely to establish *de novo* positions in those stocks.

[7] In untabulated results we run analysis for adjusted KLD score and ESG Combined Score separately. The results are mainly driven by adjusted KLD score.

From the results in Tables 7 and 8, we conclude that when investors exhibit strong interest in ESG which is proxied by high ESG reading intensity, they are more likely to invest in or less likely to sell (especially completely sell off) stocks that have better ESG performance. Besides, among the three ESG ratings we use, investors care most about KLD and Refinitiv. There is also some evidence that investors care about "reputation" of ESG risks (*Peak RRI*).

5. Assessing the joint dynamics of investor and firm ESG reading intensity.

In this section, we try to measure similarity between investor's and firm's ESG reading and identify periods in which there are increases in common reading intensity. The goal of this analysis is to assess how *commonality* in ESG reading intensity by a firm's employees and by those of the institutional investors that hold those shares in the firm influences the investor trading decisions. We use FactSet holdings data and merge with The Company's intent data as in Sections 3 and 4.

5.1 Similarity of ESG reading and investor's trading decisions

We first define how we measure the similarity in ESG topics between an investor *i* and a firm *j*. For each topic among all 323 ESG topics in any quarter, we count the number of weeks that have Spike scores for pairs of investors and firms. A cosine similarity for each investor-firm pair is computed as follows:

$$Similarity^{ESG} = \frac{\sum_1^n C_i * C_j}{\sqrt{\sum_1^n {C_i}^2} * \sqrt{\sum_1^n {C_j}^2}}$$

C_i is number of weeks for investor *i* and is C_j is number of weeks for firm *j*. The *n* represents number of topics, and, if C_i or C_j are missing, they will be put as zero. We require at least 10 non-zero inputs for both a firm and an investor when calculating cosine similarity. Similarity of non-ESG topics ($Similarity^{Not\ ESG}$) is calculated in the same way. Then, we merge similarity data with FactSet holdings data. The final sample is from the 3rd quarter of 2015 through the 4th quarter of 2018. When we only include stocks held by the investors last quarter, there are 3,295,997 investor-firm-quarter observations, among 3,768 unique investors in 4,018 US firms.

We present results in Table 9. Panel A offers summary statistics on the new *Similarity* ^{ESG} variable which ranges from zero to one and averages around 0.725 with a standard deviation of 0.211. There are reasonable levels of commonality in reading patterns among investor-firm pairs, but also noteworthy variation across them. Outcome variables for regressions in Panels B and C are the same as in Table 8. In Panel B, we use overall investment *Invest* as the outcome variable. Across all columns we add investor-quarter, stock-quarter and investor-stock fixed effects. In Model (1), we only add $Similarity^{ESG}$, the coefficient is positive and statistically significant at 1%. In Model (2), we interact $Similarity^{ESG}$ with investor's ESG reading and the result suggests that the positive relationship between similarity in ESG topics and investment only exists when investors have strong ESG interest. The coefficient of $Similarity^{ESG} \times Log(1+Spikes^{ESG})$ is positive at 0.174 (robust *t*-statistic of 3.05) while the coefficient of $Similarity^{ESG}$ becomes insignificantly negative. In Models (3) and (4), we control investor's non-ESG reading and pairwise similarity in non-ESG topics, and the results are similar, with economic significance even stronger. In Panel C, we again use different types of investment as outcome variables. Among 3,295,997 observations, there are 433,577 "Selloffs", 1,581,386 "Decreases", 745,815 "Holds," and 968,796 "Increases." The results suggest that when there is high similarity in ESG reading and strong investor ESG interest, the investor is less likely to sell or completely sell off stocks and they are more likely to hold the stocks. Heightened ESG interest, however, is not associated with greater likelihoods of increases or pick-ups in Models (4) and (5).

We interpret from these findings that investors do care about firm's alignment in ESG reading, especially when they read intensively about ESG topics or they have strong interest in ESG.

5.2 Capturing the joint dynamics of investor and firm ESG reading intensity.

As in Section 5.1, we use FactSet holdings data to merge with The Company intent data and again each investor-stock pair in FactSet holdings are used in the following quarter. Following the work by Kempf, Manconi, and Spalt (2017), we rank investors based on relative importance to both firms and investors. Specifically, at any quarter-end we first rank investors based on their dollar holdings for each firm (*Investor Rank*). This measures how important any one investor is for the firm. Then we rank firms within each investor's portfolio based on dollar holdings (*Firm Rank*), which measures the extent to which a given firm matters among the holdings of the investor. Finally, for each firm, we can secure each investor's rank by aggregating *Investor Rank* and *Firm Rank*. An investor ranked in the top five among those for a given firm we call *Top5* and those ranked outside the top five we call *rest*.

We conduct this analysis at topic-month level, and we align investor's and firm's ESG reading in the same month. At least 10 investors for each firm in any quarter is required. The results are shown in Table 10. Across all columns the dependent variable is *1{SpikeFirm}*, a dummy variable indicating whether the firm itself has a Spike score that is larger than or equal to 80 for any topic-month. These regressions include nearly 33 million firm-month-topic observations with 4,046 unique firms, 46 months (from June 2015 to March 2019), and across 323 different ESG topics. $1\{Spike^{Top5\ Inv}\}$ and $1\{Spike^{Rest\ Inv}\}$ are defined similarly for top-five investors and all-but-the-top-five investors, respectively. We add firm fixed effects and month fixed effects for all specifications. And we multiply the outcome variables by 100 to ensure interpretable coefficients.

Results for all ESG topics are in Panel A. In Model (1), we show that investor's ESG-related reading is positively associated with a firm's ESG-related reading intensity and the coefficients for both $1\{Spike^{Top5\ Inv}\}$ and $1\{Spike^{Rest\ Inv}\}$ are statistically significant at 1%. But the economic magnitudes differ. The coefficient of top 5 investors is 2.864, which is more than twice the size of that by investors outside the top 5 in rank. The economic magnitudes imply that when top 5 investors increase reading dramatically on ESG topics, there are 2.864% higher likelihood that the firm also spikes on ESG topics. Other investors are only associated with a 1.357% higher likelihood. In Models (2) and (3), we further add firm-month fixed effects as well as topic-quarter fixed effects. The results are robust to saturation with fixed effects. In Model (4) of Panel A, we interact investor's reading intensity with a firm's institutional ownership levels last quarter. We do this in order to focus our attention on the salience of ESG attention among institutional investors which typically have the greatest stakes at risk of loss. The variable *Instown* is defined as *normalized* fraction of the shares outstanding held by institutional investors. The result indicates that there is a positive relationship between an investor's and a firm's ESG reading intensity when the firm has higher institutional ownership. Again, top-five-ranked investors matter more than those outside the top five rank.

In Models (1) to (4) of Panel B, we repeat our regression tests for each of the four ESG categories: Environment, Labor, Social and Governance. Across all columns the coefficients are positive and statistically significant at 1%. The economic significance of coefficients is similar for rest investors but different for top investors. The coefficient of *1{Spike*$^{Top5\ Inv}$*}* is highest in Environment category, followed by Labor, Governance and Social.

In order to establish some internal validity to the key findings in Tables 9 and 10, we investigate a quasi-exogenous ESG event. Specifically, we study how firm attention to ESG changed around the date on which BlackRock's Larry Fink released an open letter on January 14, 2020 to CEOs, which discusses the need of firms to focus improve their ESG performance. The letter focuses on environmental issues primarily, and to a lesser extent labor issues. In Figure 1, we plot the attention of firms with the highest percentage share by Blackrock as their investor versus firms with the lowest share or no ownership by Blackrock, finding that Blackrock-owned firms read more about the environment (and to a lesser extent labor) in the next two days, relative to firms with low or zero Blackrock ownership. The top-left and top-right panels of Figure 1 illustrate these effects with the red line representing the higher reading intensity among BlackRock-owned firms and the blue line for those with firms with zero BlackRock holdings. For governance and data security issues in the bottom-left and bottom-right panels, by contrast, there is no divergence between firms heavily owned by Blackrock and those not. This experiment gives us some additional confidence of the salience of institutional ownership for ESG reading intensity as outlined in Tables 9 and 10.

6. Conclusions.

We leverage big data analytics from an unnamed firm's proprietary B2B intent data to produce a new measure of ESG attention predicated on the internet research activity of the employees of firms across the web. The analysis suggests a meaningful statistical association between firms' reading and their future ESG performance, and investors' reading and their future actions on stocks with ESG performance. In contrast to the idea that firms read about ESG issues passively in anticipation of negative news, the evidence rather supports the idea that firms and investors read intensely in order to take anticipated action that improves their ESG performance. Firm ESG ratings tend to improve across the three measures of ESG we examine, while investors seem to cut down positions in stocks with poor ESG performance. The magnitudes of these relationships are significant, in that by some measures of

ESG, a one standard deviation increase in our parsimonious measure of ESG attention is related to a 16% of a standard deviation of improvement in ESG scores.

Our analysis makes several novel contributions. First, for a literature surrounding ESG fraught with concerns about imperfections in existing ESG index-based measures, our findings suggest that these measures do in fact correspond to attention to ESG matters among employees within the firm and by investors. However, the statistical association varies across ratings, with the widely used KLD measure weakly associated with firms' actions. Interestingly, however, this appears to be the metric used the most by investors, revealed by the strength of association between KLD ratings, ESG attention and future investment actions. This suggests a potential concern with industry practice. Even more importantly perhaps, our evidence on the intensity of comovement among a firm's attention to ESG with an investor's attention to ESG provides evidence of generally difficult-to-observe interactions between firms and investors on ESG issues. Our tests suggest that the top-five owners are three times as likely to matter as others, and this influence is strongest for environmental, less so labor, issues. We believe our reading intensity measure itself can provide a valuable new tool to investors and firms in measurement on issues, like ESG, that will be of mounting importance the coming decades.

References

Albuquerque, R., Koskinen, Y., & Zhang, C. (2019). Corporate social responsibility and firm risk: Theory and empirical evidence. *Management Science*, *65*(10), 4451-4469.

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

Balasubraman, S., Peterson, R. A., & Jarvenpaa, S. L. (2002). Exploring the implications of m-commerce for markets and marketing. *Journal of the Academy of Marketing Science*, *30*(4), 348-361.

Bansal, R., Wu, D. A., & Yaron, A. (2018). Is socially responsible investing a luxury good? University of Pennsylvania working paper. *Available at SSRN 3259209*.

Barber, B. M., Morse, A., & Yasuda, A. (2020). Impact investing. *Journal of Financial Economics*, forthcoming.

Barwise, P., & Strong, C. (2002). Permission-based mobile advertising. *Journal of Interactive Marketing*, *16*(1), 14-24.

Bauer, R., Ruof, T., & Smeets, P. (2019). Get real! Individuals prefer more sustainable investments. *Available at SSRN 3287430*.

Bénabou, R., & Tirole, J. (2010). Individual and corporate social responsibility. *Economica*, *77* (305), 1-19.

Berg, F., Koelbel, J. F., & Rigobon, R. (2020). Aggregate confusion: the divergence of ESG ratings. *Available at SSRN 3438533*.

Cahan, S. F., Chen, C., Chen, L., & Nguyen, N. H. (2015). Corporate social responsibility and media coverage. *Journal of Banking & Finance*, *59*, 409-422.

Cao, J., Titman, S., Zhan, X., & Zhang, W. E. (2020). ESG Preference, Institutional Trading, and Stock Return Patterns. *Available at SSRN 3353623*.

Ceccarelli, M., Ramelli, S., & Wagner, A. F. (2019). When investors call for climate responsibility, how do mutual funds respond?. Swiss Finance Institute Research Paper #19-13.

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

Chatterji, A. K., Levine, D. I., & Toffel, M. W. (2009). How well do social ratings actually measure corporate social responsibility?. *Journal of Economics & Management Strategy*, *18*(1), 125-169.

Chava, S. (2014). Environmental externalities and cost of capital. *Management Science*, *60*(9), 2223-2247.

Chen, T., Dong, H., & Lin, C. (2020). Institutional shareholders and corporate social responsibility. *Journal of Financial Economics*, *135*(2), 483-504.

Cheng, I. H., Hong, H., & Shue, K. (2013). *Do managers do good with other people's money?* National Bureau of Economic Research working paper #w19432.

Cronqvist, H., & Yu, F. (2017). Shaped by their daughters: Executives, female socialization, and corporate social responsibility. *Journal of Financial Economics*, *126*(3), 543-562.

Dai, R., Liang, H., & Ng, L. (2020). Socially responsible corporate customers. *Journal of Financial Economics,* forthcoming.

Davidson, R. H., Dey, A., & Smith, A. J. (2019). CEO materialism and corporate social responsibility. *The Accounting Review*, *94*(1), 101-126.

DellaVigna, S., List, J. A., & Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics*, *127*(1), 1-56.

Deng, X., Kang, J. K., & Low, B. S. (2013). Corporate social responsibility and stakeholder value maximization: Evidence from mergers. *Journal of Financial Economics*, *110*(1), 87-109.

Di Giuli, A., & Kostovetsky, L. (2014). Are red or blue companies more likely to go green? Politics and corporate social responsibility. *Journal of Financial Economics*, *111*(1), 158-180.

Dimson, E., Karakaş, O., & Li, X. (2015). Active ownership. *The Review of Financial Studies*, *28*(12), 3225-3268.

Dimson, E., Karakas, O., & Li, X. (2019). Coordinated engagements. Working paper, Cambridge University. *Available at SSRN 3209072*.

Dyck, A., Lins, K. V., Roth, L., & Wagner, H. F. (2019). Do institutional investors drive corporate social responsibility? International evidence. *Journal of Financial Economics*, *131*(3), 693-714.

Dyck, A., Lins, K. V., Roth, L., Towner, M., & Wagner, H. F. (2018). *Entrenched Insiders and Corporate Sustainability (ESG): How Much Does the "G" Matter for "E" and "S" Performance Around the World?*. Working paper. University of Toronto.

Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science*, *60*(11), 2835-2857.

Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics*, *101*(3), 621-640.

Fernando, C. S., Sharfman, M. P., & Uysal, V. B. (2017). Corporate environmental policy and shareholder value: Following the smart money. *Journal of Financial and Quantitative Analysis*, *52*(5), 2023-2051.

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

Flammer, C. (2015). Does corporate social responsibility lead to superior financial performance? A regression discontinuity approach. *Management Science*, *61*(11), 2549-2568.

Gibson, R., & Krueger, P. (2018). The sustainability footprint of institutional investors. *Swiss Finance Institute Research Paper*, (17-05).

Gibson, R., Krueger, P., Riand, N., & Schmidt, P. S. (2019). ESG rating disagreement and stock returns. Swiss Institute of Finance working paper. *Available at SSRN 3433728*.

Hart, O., & Zingales, L. (2017). Companies should maximize shareholder welfare not market value. ECGI-Finance Working Paper #521.

Hartzmark, S. M., & Sussman, A. B. (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. *The Journal of Finance*, *74*(6), 2789-2837.

Heinkel, R., Kraus, A., & Zechner, J. (2001). The effect of green investment on corporate behavior. *Journal of Financial and Quantitative Analysis*, 431-449.

Hoepner, A. G., Oikonomou, I., Sautner, Z., Starks, L. T., & Zhou, X. (2018). ESG shareholder engagement and downside risk. University of Texas working paper. *Available at SSRN 2874252*.

Hong, H. G., Kubik, J. D., Liskovich, I., & Scheinkman, J. (2019). Crime, punishment and the value of corporate social responsibility. Columbia University working paper. *Available at SSRN 2492202*.

Hong, H., & Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, *93*(1), 15-36.

Jagannathan, R., Ravikumar, A., & Sammon, M. (2017). *Environmental, social, and governance criteria: why investors are paying attention*. National Bureau of Economic Research working paper #w24063.

Karpoff, J. M., Lott, Jr, J. R., & Wehrly, E. W. (2005). The reputational penalties for environmental violations: Empirical evidence. *The Journal of Law and Economics*, *48*(2), 653-675.

Kempf, E., Manconi, A., & Spalt, O. (2017). Distracted shareholders and corporate actions. *The Review of Financial Studies*, *30*(5), 1660-1695.

Kitzmueller, M., & Shimshack, J. (2012). Economic perspectives on corporate social responsibility. *Journal of Economic Literature*, *50*(1), 51-84.

Krüger, P. (2015). Corporate goodness and shareholder wealth. *Journal of Financial Economics*, *115*(2), 304-329.

Liang, H., & Renneboog, L. (2017). On the foundations of corporate social responsibility. *The Journal of Finance*, *72*(2), 853-910.

Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *The Journal of Finance*, *72*(4), 1785-1824.

Luo, H. A., & Balvers, R. J. (2017). Social screens and systematic investor boycott risk. *Journal of Financial and Quantitative Analysis*, *52*(1), 365-399.

Masulis, R. W., & Reza, S. W. (2015). Agency problems of corporate philanthropy. *The Review of Financial Studies*, 28(2), 592-636.

Matos, P. (2020). *ESG and Responsible Institutional Investing Around the World: A Critical Review*. CFA Institute Research Foundation.

Oehmke, M., & Opp, M. M. (2020). A theory of socially responsible investment. London School of Economics working paper. *Available at SSRN 3467644*.

Pastor, L., Stambaugh, R. F., & Taylor, L. A. (2019). *Sustainable investing in equilibrium* . National Bureau of Economic Research #w26549.

Pedersen, L. H., Fitzgibbons, S., & Pomorski, L. (2019). Responsible investing: The ESG-efficient frontier. Copenhagen Business School working paper. *Available at SSRN 3466417*.

Riedl, A., & Smeets, P. (2017). Why do investors hold socially responsible mutual funds?. *The Journal of Finance*, *72*(6), 2505-2550.

Schiller, C. (2018, April). Global supply-chain networks and corporate social responsibility. In *13th Annual Mid-Atlantic Research Conference in Finance (MARC) Paper*.

Servaes, H., & Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. *Management Science*, *59*(5), 1045-1061.

Starks, L. T., Venkat, P., & Zhu, Q. (2017). Corporate ESG profiles and investor horizons. University of Texas working paper. *Available at SSRN 3049943*.

Tong, S., Luo, X., & Xu, B. (2020). Personalized mobile marketing strategies. *Journal of the Academy of Marketing Science*, *48*(1), 64-78.

Figure 1: Firm Attentional Responses to BlackRock's Larry Fink letter on January 14, 2020.

In this figure, we plot the reading intensity of firms surrounding the day of January 14, 2020 on which Larry Fink of Blackrock issued the letter to CEOs regarding sustainability, emphasizing climate issues (although not exclusively). The y-axis is the percentage of total reading for firms in the top 10% of Blackrock holdings versus the firms in the bottom 10% (including those firms where Blackrock had zero holdings), normalized to be 0 at January 6. The topic clusters chosen are Environmental, Governance, Labor, and Data and Sensitive Information Protection.

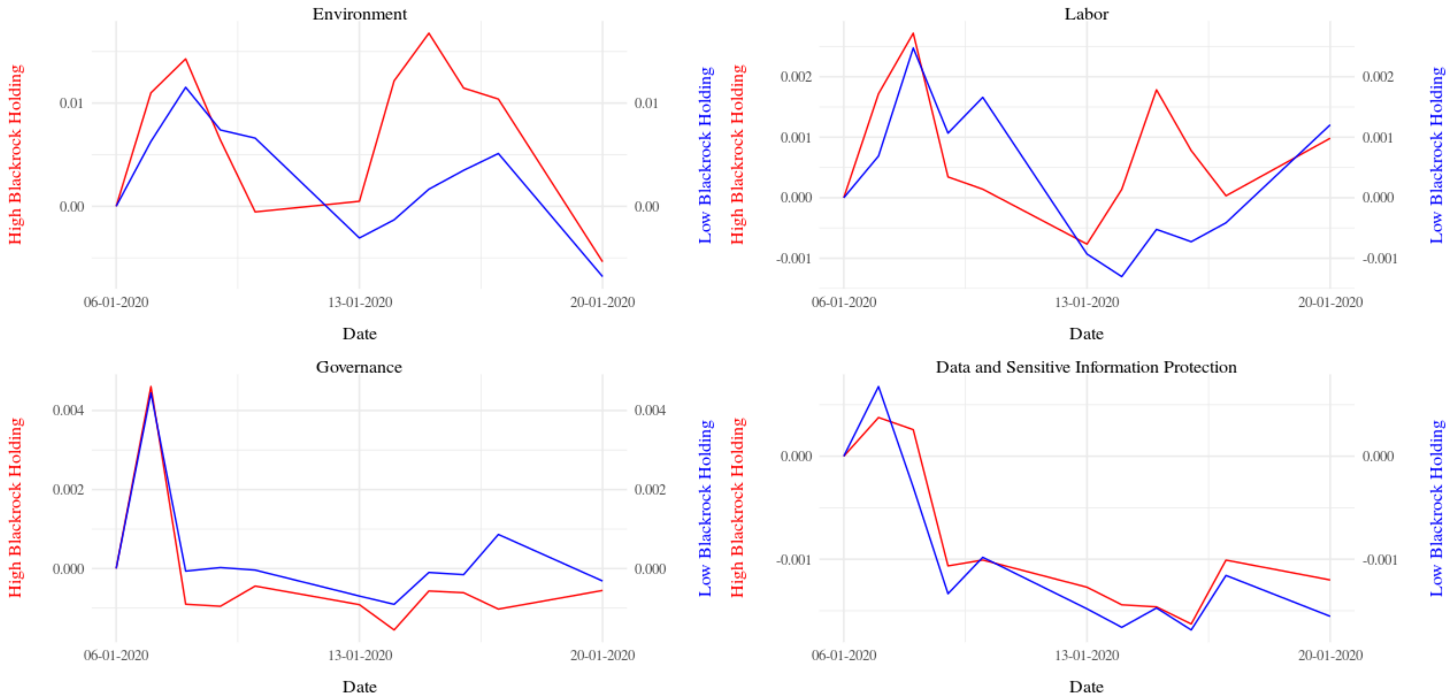

Table 1: Intent Dataset Coverage.

Panel A: Number of The Company Topics by Year

Year	# Topics	# Domains	Domain-mapped Business-Related Interactions Per Day
2015	2462	1677494	(not available)
2016	2962	1819851	506892107
2017	3589	4303994	686276353
2018	5433	5473714	623016344
2019	6765	6907293	272946594

Panel B: Number of topics in each topic category

This uses the topic taxonomy as of 2019. Themes were assigned by the data provider.

Theme name	# topics	Theme name	# topics
Events and Conferences	60	Human Resources	321
Government	87	Healthcare	327
Biotechnology	106	Energy/Construction/Manufacturing	339
Consumer Technology	149	Marketing	523
Business	301	Finance	561
Legal	305	Company	1025
		Non-consumer technology	1849

Table 2: ESG topics

Panel A: Number of topics within each ESG category

This panel shows number of topics within each ESG category we classify. There are 323 ESG topics classified to 9 categories.

Category Name	**# topics**
Compliance	18
Corporate Governance	29
CSR	28
Customer Relation	23
Cybersecurity	39
Data and Sensitive Information Protection	40
Environment	46
Equality & Diversity	21
Labor Relation	63
Total	323

Panel B: Example topics within each dimension

This panel shows 10 example topics within each 4 dimension, which we select for demonstration purpose: Environment, Labor, Compliance, and Data and Sensitive Information Protection.

Environment	**Labor**	**Compliance**	**Data and Sensitive Information Protection**
Air Pollution	Diversity Recruiting	Accounting Compliance	Data Privacy and Protection
Alternative-Fuel Vehicles	Employee Safety	Business Law	Data Security
Carbon Footprint	Employee Satisfaction	Compliance	Enterprise Application Security
Carbon Management	Equal Employment Opportunity (EEO)	Compliance Management	Information Security
Climate Change	Equal Pay / Comparable Worth	Compliance Training	Internet Security
Emissions	Gender Equality	Global Employment Law	Intrusion Prevention
Global Warming	Labor Relations	Know Your Customer (KYC)	Security Monitoring
Greenhouse Gas	Labor Union	Minimum Wage	Sensitive Data
Renewable Energy	Wellness Benefits	Regulatory Compliance	Strong Encryption
Water Pollution	Workers' Compensation	Tax Compliance	Threat Prevention

Table 3: ESG topics (continued)

Panel C: Number of The Company ESG Topics by Year

Year	# Topics	# Domains	Domain-mapped Business-Related Interactions Per Day	% of Interactions across all topics
2015	172	1520884	(not available)	(not available)
2016	188	1668113	34182462	6.74%
2017	226	4111852	64512222	9.40%
2018	323	5130072	47857872	7.68%
2019	323	6161740	20245142	7.42%

Panel D: Top Industries for Select Categories

This table shows 10 industries which have highest percentage reading within each 4 dimension: Environment, Labor, Compliance, and Data and Sensitive Information Protection. Labor includes both topics of Labor Relation and Equality and Diversity. The percentage of reading of each dimension is defined as total record of topics in that dimension divided by total record of all topics. We define industry as first two digits of NAICS code.

Environment	Labor	Compliance	Data and Sensitive Information Protection
Utilities	Educational Services	Finance and Insurance	Finance and Insurance
Mining, Quarrying, and Oil and Gas Extraction	Health Care and Social Assistance	Professional, Scientific, and Technical Services	Professional, Scientific, and Technical Services
Educational Services	Accommodation and Food Services	Administrative, Support Waste Management and Remediation Services	Health Care and Social Assistance
Construction	Management of Companies, Enterprises	Mining, Quarrying, and Oil and Gas Extraction	Accommodation and Food Services
Management of Companies, Enterprises	Administrative, Support Waste Management and Remediation Services	Accommodation and Food Services	Information
Professional, Scientific, and Technical Services	Arts, Entertainment, and Recreation	Construction	Arts, Entertainment, and Recreation
Agriculture, Forestry, Fishing and Hunting	Wholesale Trade	Educational Services	Real Estate and Rental and Leasing
Accommodation and Food Services	Agriculture, Forestry, Fishing and Hunting	Management of Companies, Enterprises	Wholesale Trade
Administrative, Support Waste Management and Remediation Services	Professional, Scientific, and Technical Services	Health Care and Social Assistance	Mining, Quarrying, and Oil and Gas Extraction
Manufacturing	Manufacturing	Real Estate and Rental and Leasing	Management of Companies, Enterprises

Table 4: Distributions of Spike Scores

Panel A All topics

This panel shows distributions of The Company's Spike Score for CRSP firms by year across all The Company topics. We report 25th, 50th, 75th, 90th, 95th and 99th percentiles of the distributions.

Year	25th	50th	75th	90th	95th	99th
2015	14	25	30	58	76	86
2016	16	25	45	60	73	85
2017	38	48	57	67	73	83
2018	40	48	57	66	71	81
2019	41	51	60	70	75	85

Panel B ESG topics only

This panel shows distributions of The Company's Spike Score for CRSP firms by year across ESG topics. We report 25th, 50th, 75th, 90th, 95th and 99th percentiles of the distributions.

Year	25th	50th	75th	90th	95th	99th
2015	14	25	35	55	71	84
2016	16	25	47	59	71	84
2017	38	48	57	66	72	83
2018	40	48	57	65	70	80
2019	41	50	60	69	74	84

Table 5: Firm's reading and ESG performance (Refinitiv)

Panel A: Summary Statistics (Annual)

Statistic	N	Mean	St. Dev.	Min	25^{th}	75^{th}	Max
$Refinitiv^{Combined}$	7582	45.265	16.238	18.317	33.540	55.520	86.039
$Refinitiv^{ESG}$	7582	50.784	18.402	19.240	35.550	66.390	89.422
$Refinitiv^{Contro}$	7582	48.252	20.020	0.000	52.580	59.090	66.670
$Log(1+Spikes^{ESG})$	7582	3.912	1.339	0.000	3.466	4.804	5.789
$Log(1+Spikes^{Environ})$	7582	2.126	1.041	0.000	1.609	2.890	3.951
$Log(1+Spikes^{Labor})$	7582	2.859	1.130	0.000	2.398	3.664	4.700
$Log(1+Spike^{Social})$	7582	1.850	1.005	0.000	1.099	2.639	4.025
$Log(1+Spikes^{Govern})$	7582	3.070	1.332	0.000	2.398	4.043	5.017
$Log(1+Spikes^{Not\ ESG})$	7582	6.484	1.663	0.000	6.019	7.543	8.769

Panel B: Level of Refinitiv ESG Score

This panel shows how firm's ESG-related reading is associated with ESG rating by Refinitiv. ESG score measures firm's relative ESG performance across more than 450 metrics, and ESG Combined Score is the ESG score with the ESG controversies overlay. For all three Refinitiv scores, higher level indicate better performance. Model (1) to 4 present results for level of ESG combined score, Model (5) and 6 show results for ESG score and ESG Controversies Score separately. We define industry by first two digits of SIC code. Standard errors are clustered on firm. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively. All continuous variables are winsorized at 1% level.

	$Refinitiv^{Combined}$				$Refinitiv^{ESG}$	$Refinitiv^{Contro}$
	(1)	(2)	(3)	(4)	(5)	(6)
$Log(1+Spikes^{ESG})$	1.240^{***}	0.304^{**}	1.108^{***}		0.180	1.703^{**}
	(0.190)	(0.137)	(0.415)		(0.205)	(0.684)
$Log(1+Spikes^{Not\ ESG})$			-0.708^{**}		-0.067	-1.206^{**}
			(0.323)		(0.163)	(0.542)
$Log(1+Spikes^{Environ})$				-0.250		
				(0.307)		
$Log(1+Spikes^{Labor})$				-0.311		
				(0.303)		
$Log(1+Spike^{Social})$				0.689^{**}		
				(0.296)		
$Log(1+Spikes^{Govern})$				0.336		
				(0.293)		
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes					
Firm FE		Yes	Yes	Yes	Yes	Yes
Observations	7582	7582	7582	7582	7582	7582
Adjusted R^2	0.077	0.696	0.696	0.696	0.946	0.431

Table 5: Firm's reading and ESG performance (RepRisk)

Panel A: Summary Statistics (Monthly)

Statistic	N	Mean	St. Dev.	Min	25th	75th	Max
Current RRI	59413	12.827	12.097	0.000	0.000	21.000	55.000
Log(1+Spikes^{ESG})	59413	1.677	1.037	0.000	1.099	2.398	4.316
Log(1+Spikes^{Environ})	59413	0.450	0.619	0.000	0.000	0.693	2.303
Log(1+Spikes^{Labor})	59413	0.876	0.821	0.000	0.000	1.386	3.497
Log(1+Spike^{Social})	59413	0.371	0.544	0.000	0.000	0.693	2.079
Log(1+Spikes^{Govern})	59413	1.080	0.954	0.000	0.000	1.792	3.466
Log(1+Spikes$^{Not\ ESG}$)	59413	4.075	1.378	0.000	3.466	4.977	6.802

Panel B: RepRisk Index

This panel shows how firm's ESG-related reading is associated with ESG-related risk by RepRisk. The outcome variable is current RRI, which measures firms' current exposure to ESG risks. In Model (1) to 4 the unit of observation is firm-month while in Model (5) to 6 the unit of observation is firm-year. We define industry by first two digits of SIC code. Standard errors are clustered on firm. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively. All continuous variables are winsorized at 1% level.

	Current RRI					
	Monthly				Annual	
	(1)	(2)	(3)	(4)	(5)	(6)
Log(1+Spikes^{ESG})	2.060***	-0.325***	-0.203***		-0.450***	-0.250
	(0.165)	(0.072)	(0.066)		(0.120)	(0.369)
Log(1+Spikes$^{Not\ ESG}$)			-0.133**			-0.178
			(0.065)			(0.316)
Log(1+Spikes^{Environ})				-0.177***		
				(0.062)		
Log(1+Spikes^{Labor})				-0.171***		
				(0.064)		
Log(1+Spike^{Social})				-0.181**		
				(0.071)		
Log(1+Spikes^{Govern})				-0.109*		
				(0.063)		
Month FE	Yes	Yes	Yes	Yes		
Year FE					Yes	Yes
Industry FE	Yes					
Firm FE		Yes	Yes	Yes	Yes	Yes
Observations	59413	59413	59413	59413	6610	6610
Adjusted R^2	0.100	0.706	0.706	0.706	0.653	0.653

Table 6: Firm's reading and ESG performance (KLD)

Panel A: Summary Statistics (Annual)

Statistic	N	Mean	St. Dev.	Min	25th	75th	Max
KLD1	7396	0.024	0.036	-0.107	0.000	0.037	0.217
KLD2	7396	0.448	0.646	-1.000	0.000	1.000	1.000
KLD3	5894	0.562	0.678	-1.000	0.273	1.000	1.000
$Log(1+Spikes^{ESG})$	7396	4.198	1.116	0.000	3.738	4.934	6.749
$Log(1+Spikes^{Environ})$	7396	2.335	0.939	0.000	1.946	2.996	4.844
$Log(1+Spikes^{Labor})$	7396	3.098	0.984	0.000	2.565	3.784	5.889
$Log(1+Spike^{Social})$	7396	2.033	0.937	0.000	1.386	2.773	4.304
$Log(1+Spikes^{Govern})$	7396	3.334	1.203	0.000	2.773	4.163	5.905
$Log(1+Spikes^{Not\ ESG})$	7396	6.842	1.284	0.000	6.353	7.640	9.541

Panel B: Adjusted KLD Score

This panel shows how firm's ESG-related reading is associated with ESG rating by KLD. Model (1) to 3 present results for the first version of KLD score, which is (Str – Con)/(n_Str + n_Con). In column 4 to 6, we focus on the intensive margin, which is defined as (Str – Con)/(Str + Con). If there are no strengths or concerns for any firm-year, the KLD score would be 0 in Model (4) but missing in Model (5) and 6. Str and Con are number of strengths and concerns the firm have in each year respectively. n_Str and n_Con are number of maximum strengths and concerns the firm could have respectively. And we multiply outcome variables by 100 in all columns. We define industry by first two digits of SIC code. Standard errors are clustered on firm. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively. All continuous variables are winsorized at 1% level.

	KLD1			KLD2	KLD3	
	(1)	(2)	(3)	(4)	(5)	(6)
$Log(1+Spikes^{ESG})$	0.751***	-0.032	-0.132	3.577	5.367**	
	(0.055)	(0.048)	(0.100)	(2.426)	(2.648)	
$Log(1+Spikes^{Not\ ESG})$			0.099	2.006	0.284	
			(0.085)	(2.304)	(2.525)	
$Log(1+Spikes^{Environ})$						1.280
						(1.923)
$Log(1+Spikes^{Labor})$						-1.197
						(2.427)
$Log(1+Spike^{Social})$						1.128
						(1.738)
$Log(1+Spikes^{Govern})$						5.041**
						(1.960)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes					
Firm FE		Yes	Yes	Yes	Yes	Yes
Observations	7396	7396	7396	7396	5894	5894
Adjusted R^2	0.146	0.790	0.790	0.612	0.689	0.690

Table 7: Investor's ESG reading and portfolio-level ESG performance

This panel shows how investor's ESG reading affect ESG performance of investor's portfolio. In Model (1) to 6, the outcome variable is equal-weighted ESG rating of stocks. Peak RRI is the maximum of RepRisk Index in the last two years, which represents past ESG reputation. In Model (7), the outcome variable is percentage of stocks that have Peak RRI which are larger than 0. In Model (7), the outcome variable is percentage of stocks that have Peak RRI which are larger than or equal to 50. Other variables are defined in Table 4,5,6. Standard errors are clustered on investor. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively. All continuous variables are winsorized at 1% level.

	KLD1	**Refinitiv^{Combined}**	**Refinitiv^{ESG}**	**Refinitiv^{Contro}**	**Current RRI**	**Peak RRI**	**% Peak RRI 0**	**% Peak RRI 50**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(1+Spikes^{ESG})	0.925**	1.268**	1.439***	-0.768	0.047	-0.621	-2.033***	-0.396
	(0.453)	(0.627)	(0.438)	(0.517)	(0.519)	(0.519)	(0.745)	(0.510)
Log(1+Spikes$^{Not\ ESG}$)	0.071	-0.465	-0.530	0.345	0.012	-0.094	0.063	-0.292
	(0.392)	(0.572)	(0.367)	(0.440)	(0.403)	(0.396)	(0.644)	(0.399)
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Investor FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31068	30813	30813	30813	30874	30874	30874	30874
Adjusted R^2	0.897	0.789	0.909	0.860	0.888	0.887	0.706	0.895

Table 8: Investor's ESG reading and trading decisions

Panel A: Summary Statistics (Quarterly)

Statistic	N	Mean	St. Dev.	Min	25th	75th	Max
Invest	6359854	-1.574	4.286	-16.308	-0.223	0.033	2.506
Log(1+Spikes^{ESG})	6359854	2.556	1.180	0.000	1.792	3.434	4.454
Log(1+Spikes$^{Not\ ESG}$)	6359854	5.073	1.445	0.000	4.369	6.100	7.327
KLD1	5878637	0.001	0.975	-2.056	-0.706	0.445	3.643
Refinitiv^{Combined}	5033633	-0.006	0.987	-1.848	-0.697	0.702	2.536
Current RRI	4891767	-0.003	0.997	-1.283	-0.882	0.382	2.959
RRI Trend (ΔRRI)	4891767	-0.001	0.946	-2.875	-0.502	0.201	3.804
ESG^{Zscore}	4796856	0.073	1.596	-2.935	-1.182	1.131	4.431

Panel B: Overall investment

This panel shows how investor's ESG-related reading and stock's ESG performance affect overall investment. The outcome variable is *Invest*, which is log change of dollar holdings adjusted by quarterly stock return (defined in Section 4.2**)**. We present results for stocks held by the investors last quarter. *Score^{ESG}* is the standardized measure of different ESG ratings. Model (1) presents results for adjusted KLD score and Model (2) presents results for ESG Combined Score. Model (3) shows results for Current RRI and Model (4) shows results for change of Current RRI (Δ*RRI*). In all columns, we multiply the outcome variable, stock return and volatility by 100. Standard errors are clustered on investor. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively. All continuous variables are winsorized at 1% level.

	Invest			
	KLD1	**Refinitiv^{Combined}**	**Current RRI**	**ΔRRI**
	(1)	(2)	(3)	(4)
Score^{ESG}	6.104***	7.948***	-3.268*	-3.368***
	(1.673)	(1.685)	(1.939)	(0.791)
Score^{ESG}×Log(1+Spikes^{ESG})	2.383***	1.390**	0.284	-0.424
	(0.638)	(0.658)	(0.788)	(0.353)
Score^{ESG} ×Log(1+Spikes$^{Not\ ESG}$)	-0.448	-0.519	1.624**	0.598**
	(0.559)	(0.586)	(0.653)	(0.296)
Return	0.302***	0.600***	0.148***	0.145***
	(0.039)	(0.037)	(0.035)	(0.035)
Volatility	-13.617***	-15.664***	-13.681***	-14.057***
	(0.863)	(0.846)	(0.777)	(0.791)
Market Capitalization	0.426***	0.401***	0.405***	0.450***
	(0.008)	(0.009)	(0.007)	(0.009)
Momentum	8.609***	10.210***	14.417***	13.037***
	(1.699)	(1.706)	(1.665)	(1.668)
Gross Profitability	-15.582***	-17.529***	-25.092***	-25.033***
	(2.307)	(2.158)	(2.278)	(2.273)
Book-to-Market	-3.333**	-6.775***	-14.251***	-13.690***
	(1.575)	(1.452)	(1.492)	(1.447)
Investor×Quarter FE	Yes	Yes	Yes	Yes
Observations	5878544	5033633	4891767	4891767
Adjusted R^2	0.322	0.363	0.317	0.316

Table 8: Investor's ESG reading and trading decisions (continued)

Panel C: Different types of investment

This panel shows different types of investment of Panel B. For demonstration purpose we use ESG^{Zscore}, which is the sum of the standardized KLD score and ESG Combined Score. Model (1) to 4 presents results for stocks held by the investors last quarter while Model (5) present results for stocks not held by the investors last quarter. In all columns the outcome variables are dummy variables and we multiply them by 100. "Selloff" is defined as selling all stocks. "Decreases" is defined as more than 1 percent decrease of dollar holdings. " Hold" is defined as within 1 percent change of dollar holdings. "Increases" is defined as more than 1 percent increase of dollar holdings. "Pickup" is defined as buying stocks not held last quarter. Standard errors are clustered on investor. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively. All continuous variables are winsorized at 1% level.

	1{Selloff}	**1{Decreases}**	**1{Hold}**	**1{Increases}**	**1{Pickup}**
	(1)	(2)	(3)	(4)	(5)
ESG^{Zscore}	-0.603***	-0.333*	0.161	0.173	-0.004
	(0.102)	(0.173)	(0.160)	(0.159)	(0.018)
$ESG^{Zscore} \times Log(1+Spikes^{ESG})$	-0.075**	0.035	-0.096	0.062	0.034***
	(0.038)	(0.073)	(0.060)	(0.067)	(0.008)
$ESG^{Zscore} \times Log(1+Spikes^{Not\ ESG})$	0.025	0.111*	-0.078	-0.033	0.018***
	(0.035)	(0.064)	(0.051)	(0.055)	(0.006)
Return	-0.044***	-0.059***	0.031***	0.028***	0.002***
	(0.003)	(0.005)	(0.003)	(0.006)	(0.0003)
Volatility	1.363***	-0.130	-0.519***	0.649***	-0.011*
	(0.067)	(0.107)	(0.081)	(0.108)	(0.006)
Market Capitalization	-0.033***	0.012***	-0.043***	0.031***	0.029***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Momentum	-0.957***	-3.312***	1.785***	1.527***	0.129***
	(0.122)	(0.280)	(0.191)	(0.245)	(0.014)
Gross Profitability	1.291***	1.246***	-1.593***	0.347	0.330***
	(0.165)	(0.226)	(0.197)	(0.247)	(0.021)
Book-to-Market	0.751***	2.349***	-1.048***	-1.301***	-0.118***
	(0.112)	(0.189)	(0.182)	(0.199)	(0.015)
Investor×Quarter FE	Yes	Yes	Yes	Yes	Yes
Observations	4796856	4796856	4796856	4796856	39241287
Adjusted R^2	0.355	0.166	0.224	0.155	0.146

Table 9: Similarity in ESG reading and investment

Panel A: Summary Statistics (Quarterly)

Statistic	N	Mean	St. Dev.	Min	25th	75th	Max
Invest	3295997	-1.645	4.350	-16.257	-0.248	0.031	2.586
$Similarity^{ESG}$	3295997	0.725	0.211	0.224	0.573	0.908	0.999
$Log(1+Spikes^{ESG})$	3295997	2.530	1.217	0.000	1.792	3.434	4.454
$Log(1+Spikes^{Not\ ESG})$	3295997	5.014	1.560	0.000	4.317	6.105	7.327

Panel B: Overall investment

This panel investigates cosine similarity in ESG topics between firms and investors and its relation to overall investment. The outcome variable is *Invest*, the log difference in positions between the prior quarter and the current quarter for holdings. We present results for stocks held by the investors last quarter. $Log(1+Spikes^{ESG})$ is investor's ESG reading. Standard errors are clustered on investor. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively. All continuous variables are winsorized at 1% level.

	Invest			
	(1)	(2)	(3)	(4)
$Similarity^{ESG}$	0.340^{***}	-0.137	0.221	-0.332
	(0.094)	(0.151)	(0.235)	(0.225)
$Similarity^{ESG} \times Log(1+Spikes^{ESG})$		0.174^{***}	0.329^{***}	0.266^{***}
		(0.057)	(0.112)	(0.093)
$Similarity^{ESG} \times Log(1+Spikes^{Not\ ESG})$			-0.151^{*}	
			(0.080)	
$Similarity^{Not\ ESG}$				0.428
				(0.312)
$Similarity^{Not\ ESG} \times Log(1+Spikes^{Not\ ESG})$				-0.091
				(0.062)
Investor×Quarter FE	Yes	Yes	Yes	Yes
Stock×Quarter FE	Yes	Yes	Yes	Yes
Investor×Stock FE	Yes	Yes	Yes	Yes
Observations	3295997	3295997	3295997	3295997
Adjusted R^2	0.537	0.537	0.537	0.537

Table 9: Similarity in ESG reading and investment (continued)

Panel C: Different types of investment

This panel shows different types of investment of Panel B. Model (1) to 4 presents results for stocks held by the investors last quarter while Model (5) present results for stocks not held by the investors last quarter. Outcome variables are the same in Table 8 and we multiply them by 100. Standard errors are clustered on investor. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively. All continuous variables are winsorized at 1% level.

	1{Selloff}	**1{Decreases}**	**1{Hold}**	**1{Increases}**	**1{Pickup}**
	(1)	(2)	(3)	(4)	(5)
$Similarity^{ESG}$	-0.685	-8.079**	7.430**	0.649	0.714***
	(1.970)	(4.108)	(3.256)	(4.307)	(0.198)
$Similarity^{ESG} \times Log(1+Spikes^{ESG})$	-2.531***	-3.983**	3.351***	0.632	-0.010
	(0.718)	(1.698)	(1.004)	(1.512)	(0.078)
$Similarity^{ESG} \times Log(1+Spikes^{Not\ ESG})$	1.010*	3.159**	-3.167***	0.008	-0.148**
	(0.575)	(1.356)	(0.755)	(1.346)	(0.062)
Investor×Quarter FE	Yes	Yes	Yes	Yes	Yes
Stock×Quarter FE	Yes	Yes	Yes	Yes	Yes
Investor×Stock FE	Yes	Yes	Yes	Yes	Yes
Observations	3295997	3295997	3295997	3295997	34479552
Adjusted R^2	0.554	0.241	0.356	0.197	0.291

Table 10: Investor's ESG reading and Firm's Reading

Panel A: All ESG topics

This panel shows how investor's ESG-related reading is associated with firm's ESG-related reading at the topic-month level. Across all columns the outcome variable is $1\{Spike^{Firm}\}$, which is a dummy variable indicating whether the firm has a Spike score that is larger than or equal to 80 for each topic-month. $1\{Spike^{Top5\ Inv}\}$ and $1\{Spike^{Rest\ Inv}\}$ are similar defined dummies for top 5 investors and rest investors of the firm. The variable *Instown* is defined as normalized fraction of the shares outstanding held by institutional investors. Standard errors clustered on firm. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively. All continuous variables are winsorized at 1% level.

	$1\{Spike^{Firm}\}$			
	(1)	(2)	(3)	(4)
$1\{Spike^{Top5\ Inv}\}$	2.864***	3.363***	1.113***	1.309***
	(0.059)	(0.059)	(0.038)	(0.039)
$1\{Spike^{Rest\ Inv}\}$	1.357***	1.659***	0.165***	0.265***
	(0.022)	(0.025)	(0.018)	(0.018)
$1\{Spike^{Top5\ Inv}\}\times Instown$				1.581***
				(0.047)
$1\{Spike^{Rest\ Inv}\}\times Instown$				0.512***
				(0.022)
Month FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Firm×Month FE		Yes	Yes	Yes
Topic×Quarter FE			Yes	Yes
Observations	32774528	32774528	32774528	32774528
Adjusted R^2	0.025	0.054	0.081	0.082

Panel B: Different ESG categories

This panel shows results for different ESG categories. Other specifications are the same as those in Panel A. Standard errors clustered on firm. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively. All continuous variables are winsorized at 1% level.

	Environment	**Labor**	**Social**	**Governance**
	(1)	(2)	(3)	(4)
$1\{Spike^{Top5\ Inv}\}$	1.741***	1.168***	0.579***	1.075***
	(0.052)	(0.041)	(0.048)	(0.046)
$1\{Spike^{Rest\ Inv}\}$	0.363***	0.080***	0.230***	0.199***
	(0.023)	(0.019)	(0.022)	(0.022)
Month FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Firm*Month FE	Yes	Yes	Yes	Yes
Topic*Quarter FE	Yes	Yes	Yes	Yes
Observations	4348245	11628818	4001277	12796188
Adjusted R^2	0.087	0.085	0.053	0.095

Appendix A. Intent Data Spikes and ESG Topics.

This appendix illustrates how we calculate the measures $Spikes^{ESG}$ (count of Spike scores of ESG topics which are at least 80) and $Spikes^{Not\ ESG}$ from the weekly Spike score in The Company's Topic Interest model. In the following example, we simplify by considering a domain-month with 4 weeks and 5 topics (2 ESG topics and 3 non-ESG topics) and show Spike score of each topic-week, but it also applies to other more general circumstances. As seen in the following table, the Spike score can be missing.

	Charitable Giving	**Climate Change**	**Innovation**	**ROA**	**Brand Loyalty**
Week1	17	51	74	**80**	NA
Week2	20	69	76	61	31
Week3	21	**83**	78	49	36
Week4	29	**85**	**81**	36	38

The two ESG topics are "Charitable Giving" and "Climate Change", and the three non-ESG topics are "Innovation", "ROA" and "Brand Loyalty". We highlight the Spike score which is larger than or equal to 80. For the topic "Climate Change", there are two weeks with Spike score that is larger than or equal to 80, so the count would be 2. Doing the same calculations to other topics and doing aggregations, $Spikes^{ESG}$ is 2 and $Spikes^{Not\ ESG}$ is 2.

Appendix B. Alternative specifications

Appendix B.1 Firm's ESG reading and change in ESG rating

This table shows how firm's ESG-related reading is associated with change in ESG rating. Model (1) presents result for annual change in ESG Combined Score by Refinitiv, Model (2) present result for monthly change in Current RRI, Model (3) to 5 show results for three versions of adjusted KLD score, which are defined in Table 6. Across all columns we control the lagged level of respective rating. Standard errors are clustered on firm. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively. All continuous variables are winsorized at 1% level.

	ΔRefinitiv^{Combined}	**ΔRRI**	**ΔKLD1**	**ΔKLD2**	**ΔKLD3**
	(1)	(2)	(3)	(4)	(5)
Log(1+Spikes^{ESG})	0.959**	-0.062**	0.001	0.046**	0.050**
	(0.418)	(0.030)	(0.001)	(0.023)	(0.023)
Log(1+Spikes$^{Not\ ESG}$)	-0.528	0.004	0.001	0.002	-0.010
	(0.331)	(0.025)	(0.001)	(0.020)	(0.020)
Lagged rating	-0.327***	-0.228***	-0.155***	-0.431***	-0.343***
	(0.013)	(0.005)	(0.010)	(0.010)	(0.013)
Month FE		Yes			
Year FE	Yes		Yes	Yes	Yes
Industry FE	Yes		Yes	Yes	Yes
Firm FE		Yes			
Observations	6530	59330	6833	6833	4597
Adjusted R^2	0.212	0.141	0.151	0.276	0.266

Appendix B.2 Firm's ESG reading and ESG rating --- Daily Aggregates

This table shows how firm's ESG-related reading is associated with ESG rating. In this table, we use daily reading counts data instead of the 'Spike' measure. Our sample period begins in May 2016, reducing our observations. Our first measure is *Log(1+%Reading^{ESG}*100)*, in which *%Reading^{ESG}* is percentage of total records that are ESG related. Our second measure is *Log(1+Reading^{ESG}/Asset)*, in which *Reading^{ESG}/Asset* is total records of ESG topics scaled by firm's total asset last year. Across all columns we control firm's reading of non-ESG topics *Log(1+Reading$^{Not\ ESG}$/Asset)* and logarithm of total assets (*Size*). Outcome variables are the same as those used in Table 4, 5 and 6. And in Models (5) and (6) we multiply outcome variables by 100. Standard errors are clustered on firm. *, **, *** indicate statistical significance at the 10%, 5% and 1% level, respectively. All continuous variables are winsorized at 1% level.

	Refinitiv^{Combined}	**Refinitiv^{Combined}**	**RRI**	**RRI**	**KLD1**	**KLD1**	**KLD2**	**KLD3**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
*Log(1+%Reading^{ESG}*100)*	1.599**		-0.124		0.002			
	(0.814)		(0.206)		(0.250)			
Log(1+Reading^{ESG}/Asset)		0.486		-0.436**		0.171**	0.057***	0.052**
		(0.507)		(0.184)		(0.081)	(0.020)	(0.023)
Log(1+Reading$^{Not\ ESG}$/Asset)	0.022	-0.333	-0.034	0.138*	0.052*	-0.039	-0.018	-0.024*
	(0.079)	(0.391)	(0.044)	(0.078)	(0.028)	(0.046)	(0.012)	(0.013)
Size	-0.883	-0.814	0.219	-0.075	0.081	0.213	0.149**	0.165**
	(1.182)	(1.189)	(0.704)	(0.714)	(0.228)	(0.233)	(0.066)	(0.076)
Month FE			Yes	Yes				
Year FE	Yes	Yes			Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE			Yes	Yes				
Observations	5851	5851	41188	41188	5424	5424	5424	4343
Adjusted R^2	0.734	0.734	0.756	0.756	0.787	0.788	0.586	0.663