Rewriting History II: The (Un)predictable Past of ESG Ratings

Finance Working Paper N° 708/2020
November 2020

Florian Berg
Massachusetts Institute of Technology

Kornelia Fabisik
Frankfurt School of Finance and Management

Zacharias Sautner
Frankfurt School of Finance and Management and ECGI

ECGI Working Paper Series in Finance

Rewriting History II: The (Un)predictable Past of ESG Ratings

Working Paper N° 708/2020
November 2020

Florian Berg
Kornelia Fabisik
Zacharias Sautner

Abstract

The explosion in ESG research has lead to a strong reliance on ESG rating providers. We document widespread changes to the historical ratings of Refinitiv ESG, formerly ASSET4, a key rating provider. Across two downloads in 2018 and 2020, we document large rewritings in ESG ratings, which are systematic and partially related to firm characteristics. The retrospective rating changes have important implications for researchers and investment professionals. Depending on whether the original or rewritten data are used, rankings and classifications of firms into ESG quantiles change. We demonstrate that these changes affect tests that relate ESG ratings to stock returns.

Keywords: ESG ratings, ESG investment, CSR

JEL Classifications: G24, G30, G34

Florian Berg
Postdoctoral Associate Economics, Finance and Accounting
MIT Sloan School of Management
Building E62-366, 100 Main Street
Cambridge , MA 04122, USA
phone: (857) 999-6680
e-mail: fberg@mit.edu

Kornelia Fabisik*
Assistant Professor of Finance
Frankfurt School of Finance and Management
Adickesallee 32-34
60322 Frankfurt am Main, Germany
phone: +49 69 154008 828
e-mail: k.fabisik@fs.de

Zacharias Sautner
Professor of Finance
Frankfurt School of Finance and Management, Finance Department
Adickesallee 32-34
60322 Frankfurt am Main, Germany
phone: +49 69 154008 755
e-mail: z.sautner@fs.de

*Corresponding Author

Rewriting History II: The (Un)predictable Past of ESG Ratings[*]

Florian Berg[a] Kornelia Fabisik[b] Zacharias Sautner[c]

November 3, 2020

Abstract

The explosion in ESG research has lead to a strong reliance on ESG rating providers. We document widespread changes to the historical ratings of Refinitiv ESG, formerly ASSET4, a key rating provider. Across two downloads in 2018 and 2020, we document large rewritings in ESG ratings, which are systematic and partially related to firm characteristics. The retrospective rating changes have important implications for researchers and investment professionals. Depending on whether the original or rewritten data are used, rankings and classifications of firms into ESG quantiles change. We demonstrate that these changes affect tests that relate ESG ratings to stock returns.

JEL classification: G24, G30, G34

Keywords: ESG ratings, ESG investment, CSR

[*]We would like to thank Rui Albuquerque, Yrjo Koskinen, Christopher Malloy, and Chendi Zhang for helpful comments. The authors have no conflicts of interest to disclose.

[a]**MIT Sloan School of Management**; E-mail: fberg@mit.edu.

[b]**Frankfurt School of Finance & Management**; E-mail: k.fabisik@fs.de.

[c]**Frankfurt School of Finance & Management and ECGI**; E-mail: z.sautner@fs.de.

1. Introduction

Research on environmental, social, and corporate governance (ESG) topics in finance, accounting, and management has exploded over the last years. The surge in academic work mirrors the massive rise in the importance of ESG principles in the investment management industry (sometimes also referred to as responsible or sustainable investing). For example, funds that invest according to ESG principles attracted net inflows of $71.1bn globally between April and June 2020, despite the Covid-19 crisis, pushing assets under management in these funds to an all-time high of over $1tn.[1] Hartzmark and Sussman (2019) document that fund flows react strongly to the ESG ratings of mutual funds, which are constructed based on the ESG ratings of their portfolio firms. In the European Union, funds will soon have the obligation to advise clients on the social and environmental aspects of their investment products. ESG issues have in turn also become a top priority for many firms, affecting their investment decisions, compensation policies, or public relations activities.

A key challenge for researchers and investment professionals lies in the measurement of a firm's "ESG quality," that is, in quantifying how well a firm performs with respect to ESG criteria (e.g., climate change, pollution, diversity, or corruption). To address this challenge, most empirical ESG analyses have resorted to ESG scores (or ratings) constructed by professional data providers. Leading vendors of such scores include MSCI, Sustainalytics, S&P Global, Vigeo-Eiris or Thomson Reuters Refinitiv ESG. The growing usage of these vendors' ESG scores has raised questions by policymakers, investors, researchers, and firms about their reliability, consistency, and overall quality. Berg et al. (2020) and Gibson et al. (2019) document large disagreement across major ESG rating providers in their evaluation of firms' ESG performance. Survey evidence by Amel-Zadeh and Serafeim (2018) shows that 82% of investment professionals use ESG information in the investment process, but 26.4% also indicate a lack of ESG rating reliability. Tang et al. (2020) show that firms connected to MSCI through institutional ownership receive higher ESG scores. Concerns were also raised in the SEC's Asset Management Advisory Committee, where members expressed the view that "ESG scores do not go back far in history and are often backfilled [...]."[2]

In this paper we document widespread changes to the historical ESG scores of Thomson Reuters Refinitiv ESG ("Refinitiv ESG" henceforth). We further show that the rewriting of

[1]See "ESG funds attract record inflows during crisis," *Financial Times*, August 10, 2020.
[2]See "ESG Subcommittee Update," *Report to the SEC Asset Management Advisory Committee*, May 27, 2020.

these scores has important implications for analyses linking ESG scores to outcome variables such as firm performance or stock returns. The ESG scores constructed by Refinitiv ESG, formerly known as ASSET4, are influential. Refinitiv ESG is a key ESG rating provider, which offers "one of the most comprehensive ESG databases in the industry" (Refinitiv, 2020a). ESG scores by Refinitiv ESG have been used (or referenced) in more than 1,000 academic articles over the past 15 years. Moreover, Refinitiv ESG data are used by major asset managers, such as BlackRock, to manage ESG investment risks.[3] Refinitiv ESG data are also referenced in an ESG white paper featured at the World Economic Forum in 2019 (WEF, 2019), and analyzed as one of the three key ratings providers in a recent OECD report (Boffo and Patalano, 2020).

To document the rewriting of the ESG scores, we downloaded at different points in time two versions of the same Refinitiv ESG data for the same set of firm-years. We downloaded the first ("initial") version of the data in September 2018, and the second ("rewritten") version two years later in September 2020. The scores that we downloaded include an overall ESG score, as well as environmental (E), social (S), and governance (G) subscores. The sample contains 29,828 firm-year observations between 2011 and 2017 from 72 countries.

After inspecting the two downloads, we observed that the ESG scores for *identical* firm-years differed between the two data versions—in some cases dramatically. In fact, not a single ESG score was the same across the two versions. Thirteen percent (13%) of the sample observations were subject to a score "upgrade," that is, the rewritten ESG score was higher than the initial ESG score. Even more remarkably, 87% of the observations were subject to a score downgrade. The data rewriting is also large economically. While the overall ESG score in the rewritten version is on average 20.6% lower than in the initial version, the percentage deviations from the initial to the rewritten version for the E, S, and G subscores are -47.4%, 8.6%, and 116.2%, respectively.[4] The rewriting in the overall ESG score affects all sample years similarly, while there is meaningful heterogeneity across years for the subscores. Notably, the retrospective downgrades in the E score were much larger in the more recent years, especially in 2016 and 2017, whereas the opposite holds true for the S score.

The differences between the two data versions raise the question of why and how the scores were changed by Refinitiv ESG. According to information by Refinitiv ESG, the score deviations originate from adjustments in its scoring methodology. This scoring adjustment came into

[3] See "BlackRock taps Thomson Reuters' ASSET4 for global ESG data," *Responsible Investor*, April 11, 2011.
[4] The large average positive G-score change is driven by outliers; the median change amounts to a small negative number (-6.9%).

effect on April 6, 2020, that is, between our two data downloads. Importantly, Refinitiv ESG applied the methodology change not just to newly created scores, but it also retrospectively modified the historical scores in its database. Though providing only general information on the methodology change, Refinitiv ESG referred to two adjustments (Refinitiv, 2020b). First, Refinitiv ESG started to take into account that not all ESG metrics feeding into the ESG scores are of equal importance to every industry. Second, while Refinitiv ESG was previously assigning a neutral score to firms which did not report on a certain metric, the new methodology assigns a score of zero to such firms. We do not take a stance on whether the revised methodology enhanced the quality of the ESG scores, that is, whether the initial or revised data are better in capturing the ESG quality of firms.

As we do not have access to Refinitiv ESG's methodology to understand and verify these changes, we use statistical methods to infer the role of different economic variables in explaining the score deviations that we observe. We demonstrate that the ex-post score changes are systematic and partially driven by reassessments of industry- and country-level drivers of ESG performance (or risks). Substantial parts of the score rewriting also play out at the individual firm level. These firm level effects can partially be explained by time-varying firm characteristics. Overall, we show that large parts of the score deviations originate from ex-post reassessments of the ESG performance of specific firms in specific years (in addition to firm fixed effects, that is, the reassessments of the performance of specific firms in general).

We then turn to the question of whether the deviations in ESG scores have implications for the estimation and interpretation of the relationship between ESG scores and outcome variables. We focus this analysis on S&P 1500 firms to control for potentially confounding factors at the country level. We first demonstrate that the ESG score deviations strongly affect ESG-based ranking of S&P 1500 firms. This in turn affects the classification of firms into different ESG quantiles. For the overall ESG score, only 68.5% of firm-year observations are classified into the top decile (top 10%) in the initial and rewritten data versions; numbers are similar for the bottom decile. The overlap is only slightly larger if we look at extreme quartiles or terciles. We find similar patterns for the classification of firms based on their E, S, and G subscores. Hence, the retrospective score rewriting leads to large changes in what are deemed to be high- or low-ESG firms. This insight is important as the classification of firms into quantiles based on ESG scores (or their subscores) is widely used in ESG research, both in asset pricing and corporate finance, and in the investment industry.[5]

We use the recent Covid-19 crisis as a setting to explore the effects of these classification changes in the Refinitiv ESG database. We thereby build on Albuquerque et al. (2020), who show that firms with higher E&S ratings *prior to* the crisis exhibited better stock market performance during the pandemic. Our objective of this exercise is *not* to replicate Albuquerque et al. (2020). Instead, we aim to show that the relationship between ESG ratings and firm performance is sensitive to whether the initial or rewritten ESG data are used.[6] Similar to Albuquerque et al. (2020), we classify firms as "high-E&S firms" if they are ranked in the top quartile of the S&P 1500 sample based on the average value of their E&S scores.[7] Our tests then compare daily abnormal returns of high- and low-E&S firms before versus after a Covid-19 event date (February 24, 2020).

Our results are remarkable. When classifying firms based on the initial E&S scores, we find no evidence that high-E&S firms performed better during the Covid-19 pandemic compared to low-E&S firms. This picture looks entirely different if we run regressions using a classification of firms based on the rewritten data. We now find strong evidence that high-E&S firms exhibited better stock market performance during the pandemic relative to low-E&S firms. Not only is the statistical significance in these regressions much elevated, but we also observe that the coefficient estimates scale up by a factor between three and ten, depending on the specification.

The large differences in results that we document have economic implications. Retrospectively, one would attribute a positive performance effect to high-E&S firms *if* one were to classify firms based on the rewritten data. However, this performance would not have been achievable with the data (or information) available to investors at the onset of (or before) the pandemic. At this point in time, investors would have classified firms differently into high- and low-E&S groups, and the performance differences between these two sets of firms would not have been economically and statistically different. Hence, the benefits of being a high-E&S firms during the crisis would have been exaggerated. The implications of this observation extend beyond

[5]For example, studies classify firms into top and bottom ESG quantiles to examine whether a portfolio that is long (short) in high-ESG stocks (low-ESG stocks) generates outperformance (Statman and Glushkov, 2009). Studies also use the classification of firms into quantiles to examine whether high-ESG firms (or high-E&S firms) performed better during crisis periods (Lins et al., 2017; Albuquerque et al., 2020).

[6]We actually differ in our estimation from Albuquerque et al. (2020): (i) we condition on E&S ratings in 2017, while they use 2018 (our sample stops in that year due to the initial 2018 download); and (ii) we examine effects for firms in the S&P 1500, while they use a broader sample. Moreover, they show that their results are robust to using a classifications based on E&S ratings by MSCI ESG.

[7]Ding et al. (2020) also use Refinitiv ESG to test how firms with high ESG scores performed during the pandemic.

our setting. They apply more broadly for the backtesting of ESG strategies, as also for such tests it is critical to verify that the original, not the rewritten, scores are being used. Of course, our insights are also critical for future ESG research using Refinitiv ESG data. A recommendation that follows from our analysis is that researchers using these data should verify whether the initial, originally-available data are needed to test their hypotheses. This consideration is important in light of the expected (continued) growth in ESG research.

Finally, we show that firms that experienced positive (or only small negative) score deviations (top quartile in the distribution of the ESG score deviation), exhibited positive announcement returns (CARs) when Refinitiv ESG's methodology change was announced. To the contrary, firms with large negative score deviations (bottom quartile) experienced negative CARs. Though we do not want to overinterpret these results, the estimated CARs are consistent with price pressure caused by changes in the portfolio allocation of some ESG investors in response to the data rewriting.[8] Alternatively, the CARs may reflect that the market gained new information about the inherent ESG quality of firms.

Our paper contributes most directly to the voluminous ESG research using data from Refinitiv ESG or ASSET4. As suggested by our title, we build in spirit on Ljungqvist et al. (2009), who document in a pioneering paper widespread data rewriting in the IBES analyst recommendations database. A difference between our and their analysis is that the rewriting of our ESG scores was made public by Refinitiv ESG, while this was not the case for the IBES rewriting. More broadly, we also relate to the large empirical asset pricing literature that has highlighted the importance of accounting for look-ahead bias when examining returns (Ter Horst and Verbeek, 2007; Baquero et al., 2005). Look-ahead bias arises whenever portfolios are constructed *ex post* using conditioning variables that were unavailable at the time of portfolio formation.

2. Data

2.1. Refinitiv ESG Scores

Our variable of interest in this paper is the Thomson Reuters Refinitiv ESG score (*ESG Score*, henceforth) as well as its environmental (E), social (S), and corporate governance (G) pillars (or subscores). The scores are constructed based on publicly-reported data and measure

[8] For example, investors that overweight firms with high-ESG scores may have sold (bought) firms experiencing large (only small) negative changes in their ESG scores to rebalance their portfolios.

a firm's ESG performance. The overall ESG scores, as well as the subscores, are percentile rank scores, and hence range between 0 (minimum score) and 100 (maximum score). According to Thomson Reuters Refinitiv ESG the scores "are based on relative performance of ESG factors with the company's sector (for E and S) and country of incorporation (for G)" (Refinitiv, 2020a). The scores were initially constructed by ASSET4, a company founded in 2003 and acquired in 2009 by Thomson Reuters. In 2017, Thomson Reuters enriched the database with thirteen new data items, with one of them being a new overall ESG score. We use this ESG score for our analysis. Thomson Reuters referred to these changes as "an enhancement and replacement to the existing equally weighted ASSET4 ratings" (Thomson Reuters, 2017).[9] The new ESG score was made available for firm-years that also had an ASSET4 rating.[10] According to Thomson Reuters, the overall ESG score is based on data metrics in ten different categories, which flow into the three E, S, and G subscores (Thomson Reuters, 2017). The E subscore thereby receives a weight of 34% in the total ESG score, the S subscore of 35.5%, and the G subscore of 30.5%. An "economic" subscore that used to be part of the initial equally-weighted ASSET4 ESG score is no longer a component part of the new ESG score. Since 2018, the ESG data of Thomson Reuters are part of Thomson Reuters Refinitiv and know as "Refinitiv ESG".

ESG scores by Refinitiv ESG are widely used, both in academic research and in the investment management industry. Figure 1 shows that, by October 2020, more than 1,200 academic articles have mentioned Refinitiv ESG data, because they use them in empirical tests or consider referencing them important. In the figure, we use the cumulative count for the search term "ASSET4," instead of Refinitiv ESG, to be able to construct a time-series of the name mentions since 2003, the year ASSET4 was founded.[11] The figure also reports the cumulative count of academic articles mentioning ASSET4 in combination with ESG. We use this refinement as some data items in ASSET4 may be used in contexts unrelated to ESG.[12] While the total count of articles is slightly lower, the massive surge in the number of articles mentioning both ASSET4 and ESG is similar.

[Figure 1 here]

[9]The earliest mention of this score that we could find dates back to March 2017, when Thomson Reuters prompted its customers to migrate from the ASSET4 score to the new ESG score (Thomson Reuters, 2017).

[10]The correlation between the old ASSET4 score (data item A4IR) and the new Refinitiv ESG score (data item TRESGS) corresponds of 0.82.

[11]Most researchers continue to mention the old name of the database after ASSET4 was acquired by Thomson Reuters. If at all, this implies that our article search understate the actual usage of the ESG scores.

[12]The acronym "ESG" was coined in 2004 by Ivo Knoepfel in a report for UN Secretary General Kofi Annan.

Table IA1 shows that ESG data from Refinitiv ESG is used in many articles published in leading finance journals. The table further shows that the data are also used in many contemporaneous working papers (we list selected papers only).

2.2. Data Downloads of Initial and Rewritten Refinitiv ESG Scores

We base our analysis on two versions of the *same* Refinitiv ESG database, which we downloaded in September 2018 and in September 2020, that is, two years apart. The first version of the database, accessed on September 25, 2018, covers the universe of firms in the database as of that date. We downloaded data for the years 2011 to 2017. We refer to this 2018 version of the Refinitiv ESG data as the "initial" version. The initial data contain 45,284 unique ISIN-year observations between 2011 and 2017, with 6,871 unique ISIN identifiers. Out of this grand total, an ESG score is available for 31,790 ISIN-year observations, corresponding to 6,636 unique ISINs. To determine the number of firm-year observations, we merged the data to Standard & Poor's Compustat-Capital IQ Global and North America database (we perform this step as a firm can have multiple ISINs). After the merge, our sample contains 29,926 firm-year observations for which an ESG score is available.

We downloaded the second version of the Refinitiv ESG database on September 29, 2020. We refer to this second database as the "rewritten" version. To construct this version, we downloaded the ESG scores for the 6,871 unique ISINs included in the initial data download.[13] After merging the data again to Compustat-Capital IQ, we obtain 30,619 firm-year observations with an ESG Score. These observations represent the 29,828 firm-year observations that are common in both data versions, plus 791 firm-year observations that were added through the 2020 version.[14]

Table 1 reports a year-by-year comparison of the number of observations in the two versions of the database.[15] Column 1 shows that the number observations common to both version of the data has gradually increased over time, from 3,244 in 2011 to 5,962 in 2017. Columns 2 and

[13] The names of the data codes were slightly changed from the initial to the rewritten version: the data item for overall ESG score was relabeled from *TRESGS* to *TR.TRESGScore*. We verified with Refinitiv ESG that both data items capture the same variable.

[14] The additions originate from firms included in the initial data version (not new firms) for which an ESG score was added in individual years with initially missing ESG scores.

[15] Formally, "years" refer to fiscal years (not calendar years) in the Refinitiv ESG/ASSET4 database (data item *clpactyear* in ASSET4 and *FinancialPeriodAbsolute* in Refinitiv ESG). The definition of a fiscal year in the Refinitiv ESG/ASSET4 data differs from Compustat, where a fiscal year corresponds to the year into which the majority of the months in the fiscal year fall. In Refinitiv ESG/ASSET4, the fiscal year is simply the year in which the firm's fiscal year ends.

3 show that ESG scores for 791 firm-year observations were added in the rewritten data version, while ESG scores corresponding to 98 firm-year observations were deleted in that version. While the deletion of observations is equally distributed across sample years, most of the ESG score additions relate to the later years of the sample, notably 2015 to 2017. Recall that, because we downloaded the initial data version in 2018, our sample periods ends in 2017.

[Table 1 here]

2.3. Deviations between the Initial and Rewritten Refinitiv ESG Scores

After downloading the data, we observed that the ESG scores for the *same* firm in the *same* year differed between the initial and rewritten data versions; in some cases dramatically. Table 2 reports different measures describing the ESG score deviations between the two data versions. We report this information for each year separately and, at the bottom of the table, also across all years. Columns 1 and 2 report the fraction of firm-year observations that experienced an ESG score down- or upgrade in the 2020 version. The two columns reveal a striking observation: no single ESG score is the same across the two versions of the data. Thirteen percent (13%) of the sample observations were subject to a score "upgrade," that is, the ESG score in the rewritten version ($ESG\ Score^{Rewritten\ Data}$) is higher than the ESG score in the initial version ($ESG\ Score^{Initial\ Data}$). Even more remarkably, 87% of the sample observations were subject to a score downgrade, that is, the ESG score in the rewritten version is lower than the ESG score in the initial version. The table further shows that the frequency of ESG score up- and downgrades is largely unrelated to any specific years of the sample.

In Columns 3 to 10 we report mean and median values of a measure of the magnitudes of the ESG score changes, calculated as the percentage score deviations between the initial and rewritten data versions:

$$\Delta\, ESG\ Score = \left(\frac{ESG\ Score^{Rewritten\ Data}}{ESG\ Score^{Initial\ Data}} - 1 \right) \times 100. \tag{1}$$

We report these "relative score deviations" in Columns 3 and 4 for the overall ESG score, and in Columns 5 to 10 for the three E, S, and G subscores. We calculate the deviations only for firm-year observations common to both data sets.[16] Across the full sample, Table 2 shows that

[16] Table IA2 reports absolute score deviations ($ESG\ Score^{Rewritten\ Data}$ minus $ESG\ Score^{Initial\ Data}$).

the mean score deviation for *ESG Score* amounts to -20.6%.[17] Interestingly, the mean percentage deviations across the E, S, and G subscores differ substantially, amounting to -47.4%, 8.6%, and 116.2% across the sample for Δ *E Score*, Δ *S Score*, and Δ *G Score*, respectively. Hence, the rewriting of the Refinitiv ESG data caused only modest relative changes for firms' S scores, but very large deviations for their E scores. The large average positive changes for the G score seem to be driven by a few outliers experiencing strong score upgrades, as the median changes are a small negative number (-6.9%).

[Table 2 here]

A closer look across the individual sample years reveals that the rewriting of the overall ESG score affected all years similarly, with the average values for Δ *ESG Score* ranging between -18.9% and -21.5% per year. These modest year-by-year differences for the total score mask again meaningful heterogeneity for the three subscores. Notably, the negative median changes in the *E Score* were much larger in the more recent years, especially in 2016 and 2017, compared to the earlier years, whereas the opposite holds true for the *S Score*. The median *G Score* changes seem to mostly originate from the second half of the sample.[18] We will explore the drivers of the heterogeneity across the three subscores below. We will also document that the deviations between the two versions of the data substantially altered the ESG ranking of firms, with important implications for empirical tests using these scores.

Table 3 presents further summary statistics for the initial and rewritten Refinitiv ESG data. While the average (median) ESG score in the initial version equals 50.4 (49.9), the average (median) score in the rewritten data is only 41.8 (39.9). Turning to the three subscores, we see that the average *E Score* in the initial version is 50.4, while the corresponding rewritten score is 32.8. The average *S Score* corresponds to 50.5 in the initial data, falling to 42.1 in the rewritten data. In comparison, the drop in the *G Score* is lower; it drops from an average of 50.1 in the initial version to 48.5 in the rewritten version.

[Table 3 here]

[17]There are four observations for which we cannot compute the relative change because the ESG score in the initial data is zero.

[18]These conclusions remain unchanged if we consider the absolute score deviations in Table IA2.

2.4. Methodology Changes to Refinitiv ESG Scores

The remarkable differences between the two versions of the Refinitiv ESG data raise the question of why and how the scores were changed *ex post*. According to Refinitiv ESG, the score differences originate from a change in their ESG scoring methodology, which came into effect on April 6, 2020, that is, between our first and second data downloads. Crucially, according to Refinitiv ESG and consistent with Table 2, the recalculation of the scores was not just applied to newly created ESG scores, but it also affected historical scores.[19] The public was notified of the changes in a press release on April 15, 2020 (Refinitiv, 2020b).[20]

Before exploring the implications of these score deviations, we examine more closely the reasons provided by Refinitiv ESG to explain their ESG score changes. In a document released on April 15, 2020, two main changes were highlighted: (1) changes to a materiality matrix; and (2) changes in transparency/investment grade scores (Refinitiv, 2020b). Relying on information provided by Refinitiv ESG, we summarize these changes below (see Table IA3 for details):[21]

(1) *Materiality matrix.* This change arose from a newly-introduced proprietary model that takes into account that not all ESG metrics making up the ESG scores have the same importance to every industry. A new proprietary "magnitude matrix" assesses the materiality of different metrics and assigns industry-specific weights.

(2) *Transparency/investment grade scores.* This change led to a changes in the treatment of unreported items. Previously, a score of 0.5 was allocated to companies which did not report information on a metrics, essentially giving them the "benefit of the doubt." To encourage disclosure and transparency, the changed methodology now assigns to such companies a score of zero.

We do not take a stance on whether the revised methodology enhanced the quality of the ESG scores. Hence, we do not aim at giving guidance on which of the two data versions is preferred when capturing a firm's ESG quality. Instead, our objective is to demonstrate the implications of the resultant data rewriting.

[19] This information was confirmed to us on October 5, 2020 by Refinitiv's ESG Content Support Representative.

[20] According to Refinitiv, investors subscribing to the data were notified of this change on March 6, 2020.

[21] There was also a change in how firm size affects the calculation of the "ESG Controversies Score." Since we do not analyze this score, this change does not impact our analyses.

2.5. Firm Characteristics

Summary statistics of financial characteristics our sample firms are reported in Table 3. The average ratio of capital expenditures to assets, *Capex/Assets*, is 5%, sample firms hold on average 15% of the value of their book assets in cash (*Cash/Assets*), and the average book leverage equals 24% (*Debt/Assets*). The average firms' profitability, measured as EBIT over assets, equals 7% (*EBIT/Assets*), and the average firms' property, plant, and equipment in relation to their assets correspond to 27%. R&D over assets is on average 2% (*R&D/Assets*), and the sample firms experience an average (median) sales growth of 9% (5%) (*Sales Growth*).

The distribution of the firm-year observations across countries and industries is reported in Table IA4 and Table IA5.[22]

3. Empirical Results

3.1. Economic Drivers of the Deviations in ESG Scores between the Initial and Rewritten Data Versions

We next examine more closely the role of different economic variables in explaining the deviations in ESG scores between the initial and rewritten data versions. The objective of this exercise is to document that the ex-post score changes are systematic, partially driven by reassessments of industry and country level drivers of ESG risks, and related to firm characteristics. Furthermore, the analysis will show that substantial parts of the score rewriting play out at the individual firm level. We start the analysis by conducting an analysis of variance for the ESG score deviations between the initial and rewritten versions. We then relate the deviations to firm characteristics.

3.1.1. Variance Decomposition of the Deviations in ESG Scores

Table 4, Panel A, reports the incremental explanatory power from conditioning each of our measures of the deviations in ESG scores between the initial and rewritten data on various sets of fixed effects. Time fixed effects, i.e., economy-wide reassessments of ESG aspects in specific years that affect all firms in all countries, explain very little of the variation—yielding an incremental R^2 below 2% for $\Delta\, ESG\, Score$ and all three subscores. For industry fixed effects,

[22] To classify firms into industries, we use the Fama and French 49 industry classification.

a similar observation holds true with respect to the variation in $\Delta\, ESG\ Score$, $\Delta\, E\ Score$, and $\Delta\, G\ Score$ (6.6%, 6.2%, and 3.2%, respectively). To the contrary, $\Delta\, S\ Score$ has a sizeable industry component (20.1%), which likely stems from a revised assessment of the ESG risks of specific industry sectors. The interaction of industry and time fixed effects accounts for, at most, an additional 3% of variation (in the case of $\Delta\, S\ Score$). We find only modest additional explanatory power for $\Delta\, ESG\ Score$, $\Delta\, E\ Score$, and $\Delta\, S\ Score$ when we include country fixed effects (below 10%). However, country fixed effects explain a substantial part of the variation for $\Delta\, G\ Score$, indicating that these changes can to some extent be explained by factors at the country level. Hence, meaningful parts of the reassessment of firms' governance risks originate from a reassessment of the quality of corporate governance rules and regulations across countries. Interestingly, depending on the specific measure, between 51.1% and 86.5% of variation in the score changes is *unexplained* by these sets of fixed effects, which means that variation plays out at the firm level rather than at the level of the country, industry, or over-time. Following Hassan et al. (2019), we refer to this within-country and industry-time variation as "firm level."

[Table 4 here]

In Table 4, Panel B, we decompose the firm-level variation by adding firm fixed effects. We find that permanent differences across firms in an industry and country account for 62.5%, 65.9%, 53.9%, and 37.5% of variation of $\Delta\, ESG\ Score$, $\Delta\, E\ Score$, $\Delta\, S\ Score$, and $\Delta\, G\ Score$, respectively. The remaining 37.5%, 34.1%, 46.1%, and 62.5% come from variation over time in the identity of firms in industries and countries most affected by the respective divergence in the score measures. In other words, substantial parts of the score changes originate from ex-post reassessments of the ESG performance of specific firms in specific years (on top of the reassessments of the performance of specific firms in general, as captured by the firm fixed effects).

3.1.2. Deviations in ESG Scores and Firm Characteristics

Having documented economically meaningful variation at the firm level for the score deviation measures, we next examine their correlations with a series of fundamental firm characteristics. We perform this analysis to examine the extent to which the ESG score rewriting can be explained with firm financials that vary by firm and over time. Our specification isolates the "firm-level" variation by including a full set of fixed effects (i.e., industry-by-time and country). For each firm i and year t, the empirical model is:

$$\Delta ESG\ Score_{it}^{T} = \gamma X_{it} + \delta_c + \delta_j \times \delta_t + \epsilon_{it} \tag{2}$$

where $\Delta\ ESG\ Score_{it}^{T}$ is either $\Delta\ ESG\ Score_{it}$, $\Delta\ E\ Score_{it}$, $\Delta\ S\ Score_{it}$ or $\Delta\ G\ Score_{it}$, and where the vector X_{it} contains a set of firm characteristics that includes $Log(Assets)$, $Sales\ Growth$, $Capex/Assets$, $Cash/Assets$, $Debt/Assets$, $EBIT/Assets$, $PPE/Assets$, and $R\&D/Assets$. In one specification we also add $ESG\ Score^{Initial\ Data}$ to the covariates to examine whether the score changes depend on the level of a firm's ESG score. δ_c, δ_j, and δ_t represent country, industry, and time fixed effects, respectively.

[Table 5 here]

Table 5 presents Ordinary Least Squares estimates of Equation 2; t-statistics based on standard errors clustered at the firm level are reported in parentheses. Column 1 shows that larger firms (as measured by total assets) experience an improvement in the overall ESG score in the rewritten data version. Column 3 shows that this overall increase at larger firms originates mostly from a better E score. In fact, the overall increase in the total ESG score masks that both the S score (Column 4) and the G score (Column 5) of larger firms decline in the rewritten data. There is similar heterogeneity with respect to sales growth. While firms with higher sales growth experience negative score deviations (downgrades) for the total ESG score and the E score, they experience upgrades for their S and G scores. The economic magnitudes of the effects are meaningful. A one-standard deviation increase in $Sales\ Growth$ (0.29) is associated with a decrease in $\Delta\ ESG\ Score$ of 1.5, or about 8% of the variable's standard deviation.

We further find that firms that invest more experience a reduction in the overall ESG score, originating from a downgrade in the G score. Firms that are more profitable (measured using $EBIT/Assets$) and spend more on R&D also experience upgrades in their rewritten ESG scores. However, there is some divergence across these two variables regarding the origin of the changes; while the overall improvements for more profitable firms originate from higher E and G scores (with S scores being lowered for them), the overall improvements for more R&D-intense firms are generated from upgrades in their E and S scores (with G scores being lowered for them). Finally, the estimate in Column 2 reveals that firms with higher ESG scores in the initial data experience stronger ESG score increases from the initial to the rewritten versions.

The bottom line of Table 5 is that the deviations in ESG scores between the initial and rewritten data are partially systematic, by being related to observable firm characteristics. Interestingly, these firm characteristics do not have uniform effects across all three ESG subscores; some characteristics have a positive effect on one subscore, but a negative effect on another subscore. Yet, across all specifications, we find that the overall adjusted R^2s are below 50%, that is large parts of the variation remain unexplained by the firm financials and the included fixed effects.

3.2. Economic Implications of the Deviations in ESG Scores between the Initial and Rewritten Data Versions

The previous sections document a large and partially systematic rewriting of the ESG scores in the Refinitiv ESG database. We now turn to the question of whether these changes have significant implications for researchers and investors in terms of how they estimate and interpret the relationship between ESG scores and outcome variables. To document these implications, we focus on firms in the S&P 1500—this allows us to hold the institutional setting constant and to control for potentially confounding factors at the country level (this is relevant for our Covid-19 analysis). We first show that the deviations in ESG scores caused by the rewriting of the data strongly affects ESG-based firm rankings, with important implications for the frequently used classification of firms into ESG quantiles. We then document that the relationship between these ESG quantiles and firm performance differs strongly depending on whether a classification based on the initial or rewritten data is employed.

3.2.1. Quantile Overlaps between the Initial and Rewritten ESG Data Versions

To illustrate changes in the ranking of firms, we calculate for each S&P 1500 firm with nonmissing ESG data the change in the ESG rank from the 2018 (initial) to 2020 (rewritten) version of the data. Figure 2 describes the distribution of these rank changes by reporting the fraction of firms subject to different decile rank transitions.[23] A value of 0 indicates no change in a firm's decile rank, while a decile rank change of five indicates that a firm's ranking moved

[23]We determine each June which firms are constituents of the S&P 1500 and use their ESG scores from the past fiscal year to calculate quantiles. For example, in June 2012, we keep the firms in the S&P 1500 as of that date and use the ESG scores from fiscal year 2011 as reported in the initial and rewritten data to calculate the decile ranks.

by five deciles.

[Figure 2 here]

Figure 2 shows that in more than 60% of the firm-year observations a transition in the decile rank occurred. Most of the rank changes reflect moves from one decile to the next, but a substantial number of firms also experienced changes by two, three, or four deciles. In only 37.4% of firm-year observations no decile change occurred as a result of the data rewriting.

Figure 3 expands this analysis by documenting the decile rank changes across individual sample years. As is apparent from the figure, the rank changes occurred in all sample years, with the most extreme rank changes materializing in the later years of the sample.

[Figure 3 here]

Building on these two figures, Table 6 reports the fraction of firm-year observations (in %) assigned to various extreme quantiles in the initial *and* the rewritten ESG data. Specifically, we display results for the top and bottom deciles (10%), quartiles (25%), and terciles (33%). Panel A shows the quantile overlaps for the ESG score, while Panels B, C, and D show the quantile overlaps for the E, S, and G subscores, respectively. Panel E shows the same statistics for an E&S subscore, which we calculate as the average of a firm's E and S subscores (we employ this E&S classification in the subsequent tests).[24]

[Table 6 here]

The numbers in Table 6 are striking. Across all five panels, we find only modest overlaps in the classification of firms into the different quantiles. In Panel A for the *ESG Score*, only 68.5% of S&P 1500 firm-year observations are classified into the top decile in both the initial and the rewritten data version. We find similar numbers when considering the bottom decile of the ESG score: only 62.4% of firm-year observations are common to both bottom deciles. The overlap is somewhat larger if we consider extreme quartiles or terciles, instead of extreme deciles, but substantial differences remain. For example, only 82.7% of firm-year observations are common to the top-ESG quartile in both the initial and rewritten data. We find similar

[24] Table IA6 provides the same classification using observations that are in *either* the initial *or* in the rewritten data versions. The results are similar.

patterns if we consider the E, S, and G subscores in Panels B to D, as well as the combined E&S score in Panel E. When considering the top quartiles of the subscores, the overlap across the two data versions is smallest for the $G\,Score$ in Panel D (only 55.8%), followed by the $S\,Score$ in Panel C (71.8%), and the $E\,Score$ in Panel B (74.4%). The next subsection demonstrates that this divergence in the classification of "high-ESG firm" has important implications for empirical research.

3.2.2. Effect of Initial and Rewritten E&S Scores on Stock Returns during the Covid-19 Pandemic

The classification of firms into deciles, quartiles, or terciles based on ESG scores (or subscores) is widely used in empirical ESG research, both in asset pricing and corporate finance. Industry studies by asset managers that aim at corroborating the success of certain ESG investment strategies also frequently use such categorizations. For example, studies often classify firms into top- and bottom-ESG quantiles to examine whether a portfolio that is long in high-ESG stocks, and short in low-ESG stocks, generates outperformance (Statman and Glushkov, 2009). Other studies use the classification of firms into ESG quantiles to examine whether high-ESG firms performed better during certain periods of time. For example, Lins et al. (2017) study how high-ESG firms performed during the financial crisis, and Albuquerque et al. (2020) examine how high-E&S firms performed during the Covid-19 pandemic.

We use Covid-19 as a setting to explore the effects of the changes in the ESG classification of firms induced by the rewriting of the Refinitiv ESG database. We thereby build on recent research by Albuquerque et al. (2020), who document that firms with higher E&S ratings *prior to* the crisis exhibited better stock market performance during the first quarter of 2020. Albuquerque et al. (2020) use ESG data by Refinitiv ESG to classify firms into high and low E&S groups. We want to stress that the objective of this exercise is *not* to replicate Albuquerque et al. (2020). Instead, we aim to build on their analysis to show that the relationship between ESG ratings and firm performance is sensitive to whether the initial or rewritten ESG data are used. In fact, our estimation differs from Albuquerque et al. (2020) in several dimensions: (i) we condition on E&S ratings in 2017 (they use 2018); and (ii) we examine effects for firms in the S&P 1500 (they use a broader sample).[25]

[25]Moreover, while using the Refinitiv ESG for their main tests, Albuquerque et al. (2020) also examine the effects of a classifications based on ratings by MSCI ESG. In these tests they continue to find that firms with better E&S ratings exhibited better performance during the pandemic.

Similar to the difference-in-differences (DiD) specification in Albuquerque et al. (2020), we compare daily abnormal stock returns of high- and low-E&S firms during the first quarter of 2020 before versus after a Covid-19 event date (February 24, 2020). Similar to Albuquerque et al. (2020) we classify firms as high-E&S firms if they are ranked in the top quartile of our S&P 1500 sample based on their average E&S scores. For each firm i and day t, we estimate the following empirical model:

$$\begin{aligned} Abnormal\ Returns_{it} = \beta_1 ES\ Treatment_i \times Post\text{-}COVID_t + \\ \beta_2 ES\ Treatment_i + \beta_3 Post\text{-}COVID_t + \delta_i + \delta_t + \epsilon_{it} \end{aligned} \tag{3}$$

where $Abnormal\ Return_{it}$ is the daily abnormal return of a stock, calculated as the daily log-return minus the stock's CAPM beta times the daily log-return of the market. The CAPM beta is estimated by using daily returns from 2017 to 2019 and the market return is proxied using the S&P 500.[26] *ES Treatment* equals one for firms ranked in the top quartile based on the E&S score in 2017, and zero otherwise. Importantly, *ES Treatment* is constructed using either the 2018 (initial) or the 2020 (rewritten) versions of the ESG data. *Post-COVID* equals one from February 24, 2020 to March 31, 2020, and zero before this period. Our definition of when the Covid-19 shock started to affect firms follows Albuquerque et al. (2020). As in their work, in some specifications we include the interaction of *Post-Fiscal* times *ES Treatment*, to account for the effects of the second Coronavirus Emergency Aid Package, which started the strong fiscal and monetary policy response to the pandemic in the U.S. *Post-Fiscal* equals one from March 18, 2020 to March 31, 2020, and zero before this period (the aid package was signed by president Trump on March 18, 2020). Finally, δ_i and δ_t are firm and day fixed effects, respectively. t-statistics, based on standard errors clustered by firm and day, are reported in parentheses.

[Table 7 here]

Table 7 reports the results of the DiD regressions. We report in Columns 1 and 2 regression estimates using the initial version of the E&S data, and in Columns 3 and 4 estimates with the rewritten data. Panel A reports the baseline results, while Panel B additionally controls

[26] The calculation of abnormal returns follows Albuquerque et al. (2020). Results do not differ if we use excess returns calculated as the log of 1 plus the return minus the risk-free rate. The broader sample used by Albuquerque et al. (2020) amounts to 134,689 firm-day observations in comparison to 81,163 S&P 1500 firm-day observations in our paper. One singleton observation is dropped in analyses with fixed effects.

for the Coronavirus Emergency Aid Package. In Panel A, we find in Columns 1 and 2 no statistically significant evidence that high-E&S firms exhibited better stock market performance from February 24 to March 17 compared to low-E&S firms. These results hold independently of whether we control for firm and day fixed effects. The coefficient estimates on $ES\,Treatment_i$ $\times$Post-COVID$_t$ are positive with t-statistics of around on 1.4. The size of both the coefficient estimates and the t-statistics decrease substantially in Panel B when controlling for the stimulus package (t-statistics below 0.4).

Importantly, the picture looks entirely different if we run our regressions in Columns 3 and 4 using the classification of firms into the high-E&S group based on the rewritten data. In Panel A, we now find strong evidence that high-E&S firms exhibited better stock market performance relative to other firms from February 24 to March 17. Not only is the statistical significance in these regressions much higher, with t-statistics of 2.36 and 2.37, respectively, but we also observe that the coefficient estimates are larger by a factor of about three. In Column 3 we find that high-E&S firms earned a cumulative abnormal return that is 1.3% (16 x 0.084%) higher than that of low-E&S firms. This effect is robust to the inclusion of firm and day fixed effects. The economic magnitude of the estimated effect increases once we control in Panel B for the fiscal stimulus. Notably, the magnitudes of the estimated effects using the rewritten data versions are now larger by a factor of ten compared to the estimates using the initial date (the statistical significance is lower in Columns 3 and 4 of Panel B compared to the same columns in Panel A).

The regression estimates in both panels for *Post-COVID* are similar in magnitude and statistical significance across the specifications with the two data versions. This finding is plausible: the reclassification of firms induced by the ESG data rewriting should not affect this variable. Similar conclusions follow for the estimates on *Post-Fiscal* in Panel B. This is yet again much different for the estimates on *ES Treatment* in the two data versions; while the coefficients are all negative, their magnitudes differ largely depending on whether the initial (less negative returns) or rewritten (more negative returns) data are used.

The large differences that we document in Table 7 have implications for researchers and investors. Retrospectively, one would attribute a positive performance effect during the Covid-19 pandemic to high-E&S firms *if* one were to classify firms based on the rewritten data. However, this performance would not have been achievable using the data (or information) available to investors at the onset of (or before) the pandemic. At this point in time, investors would have classified firms differently into high- and low-E&S groups, and the performance differences between these two groups over the subsequent weeks of the pandemic would not have been different from each other. Hence, the benefits of being a high-E&S firms in the crisis would have been exaggerated.

3.2.3. Event Study Returns around Refinitiv ESG's Methodology Change

We complement our analysis with a test of the stock market reaction of S&P 1500 firms around the day in which the scoring methodology change was communicated by Refinitiv ES to investors. On March 6, 2020, Refinitiv ESG issued a "content change notification" announcing the new methodology. Though it became effective only one month later, on April 6, 2020, we use March 6, 2020 as the event date because the methodology change descriptions made reasonably clear which firms would likely benefit or suffer the most from the change. For example, the fact that the nonreporting of certain ESG metrics will penalize firms allowed investors to gauge which set of firms are likely to suffer from the new methodology.

To measure the market reaction to the methodology change announcement, we assign firms to two portfolios based on the relative ESG score deviation between the initial and the rewritten data. The first portfolio is composed of firms in the top quartile of the score deviation, and the second portfolio of firms in the bottom quartile.[27] For each portfolio, we then calculate the daily equal-weighted portfolio return r_{pt} as:

$$r_{pt} = \frac{1}{N_{pt}} \sum_{i=1}^{N_{pt}} r_{ipt} \tag{4}$$

where N_{pt} is the number of firms in portfolio p at time t, and r_{ipt} is the return on firm i in portfolio p on day t. Using a four-factor model of the three Fama-French factors (Fama and French, 1993) and a momentum factor (Carhart, 1997), we then estimate the following:

$$r_{pt} - r_{ft} = \alpha_p + \beta_{p1}(r_{mt} - r_{ft}) + \beta_{p2} HML_t + \beta_{p3} SMB_t + \beta_{p4} UMD_t + AR_p \times d_t + \epsilon_{pt} \tag{5}$$

where r_{pt} is the daily equal-weighted portfolio return, r_{ft} is the return on the risk-free asset (one-month Treasury bill rate), where r_{mt} is the CRSP value-weighted return on all NYSE, NYSE MKT, Nasdaq, and Arca stocks, HML, SMB, and UMD are the value, size, and momentum factors, and ϵ_{pt} is the error term. Importantly, d_t is an indicator variable that equals one during

[27]We use data from 2017 to classify firms as we do not have the latest (2019) ESG score data before the announcement. However, the ESG scores are highly correlated over time. For example, the 2016 and 2017 ESG scores exhibit a correlation of 0.95. If at all, using the 2017 biases against finding any differences across firms.

the event window, and zero otherwise. We use three event windows: a one-day window [0;0], a two-day window including the day before the event [-1;0], and a three-day window including the day after the event [-1;1]. We capture "normal" returns by estimating Equation 5 over the 250 trading days before the event windows. The cumulative abnormal return (CAR) for portfolio p is then calculated as $T \times AR_p$, where T corresponds to the number of days of the event window.

[Table 8 here]

Table 8 reports the result for the two portfolios across the three event windows. For the announcement day [0;0] we find a positive stock price reaction for firms whose score deviation is in the top quartile, that is, for firms that saw relatively favorable ESG score changes.[28] The announcement-day CAR for these firms equals 0.41%. To the contrary, firms in the bottom quartile saw a negative CAR of -0.52%. The CAR for a long-short portfolio of the top-quartile minus the bottom-quartile group is economically sizeable and statistically significant with a return of 0.93% on the announcement days. This return is even larger if we extend the event window to [-1;0] or [-1;1] days.

Though we do not want to overinterpret these event-study results, the CARs that we document are consistent with changes in the portfolio allocation of some ESG investors. For example, investors that overweight firms with high-ESG scores may have sold (bought) firms experiencing large negative changes (increases/small negative changes) in their ESG scores. Alternatively, the CARs may reflect that the investors obtained new information about the inherent ESG quality of firms.

4. Conclusion

The explosion in ESG research has led to a strong reliance on ESG rating providers. These data vendors develop scores that evaluate how well a firm performs with respect to various ESG criteria. In this paper, we document widespread changes to the historical ESG scores of Thomson Reuters Refinitiv ESG ("Refinitiv ESG"), a key rating provider. Across two data

[28]Note that most firms experienced a decline in the ESG score as a result of the new methodology. The top-quartile portfolio therefore contains firms with positive ESG score deviations (upgrades) and some firms with modest declines in their scores. The two portfolios therefore essentially capture the *relative* strength of the changes. Firms with small negative score deviations in the top quartile may still benefit from the rewriting because they did "less poorly" compared to firms in the bottom quartile.

downloads in September 2018 and September 2020, we observe ESG rating changes for identical firm-years. The changes that we document are due to a modification in the score calculation methodology introduced by Refinitiv ESG in April 2020. Notably, Refinitiv ESG applied this modification not just to new ESG scores, but it also rewrote the historic ESG scores of firms in their database.

The rating deviations between the two downloads are systematic and partially driven by reassessments of industry- and country-level drivers of ESG risks. Substantial parts of the score rewriting also play out at the individual firm level. We demonstrate that the retrospective score changes have important implications for empirical research. Depending on whether the original or rewritten data are used, firm rankings and frequently used classifications of firms into ESG quantiles change. These changes affect tests that relate ESG ratings to firm performance or stock returns. We demonstrate these effects using the Covid-19 pandemic as a laboratory. We thereby build on Albuquerque et al. (2020) who show that firms with higher Refinitiv E&S ratings prior to Covid-19 exhibited better stock market performance during the pandemic. When we classify firms based on ESG scores in the initial data download, we find no evidence that high-E&S firms performed better during Covid-19. However, if we run regressions using a classification based on the rewritten data, we find strong evidence that high-E&S firms exhibited outperformance relative to other firms. The coefficient estimates are different by a factor of three to ten, depending on the specification.

The large differences in results that we document have meaningful implications for empirical test strategies using Refinitiv ESG data. Moving forward, researchers and investment professionals need to verify whether the original, not the rewritten, ESG scores are needed to perform their tests. Given that ESG research and ESG-related investment strategies are likely to grow even further, this is an important caveat for the use of the current, and thereby rewritten, Refinitiv ESG data.

References

Albuquerque, R. A., Y. J. Koskinen, S. Yang, and C. Zhang (2020). Resiliency of environmental and social stocks: An analysis of the exogenous COVID-19 market crash. *Review of Corporate Finance Studies 9*(3), 593–621.

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

Bae, K.-H., S. E. Ghoul, O. Guedhami, and X. Zheng (2020). Board reforms and dividend policy: International evidence. *Journal of Financial and Quantitative Analysis*. Forthcoming.

Baquero, G., J. Ter Horst, and M. Verbeek (2005). Survival, look-ahead bias, and persistence in hedge fund performance. *The Journal of Financial and Quantitative Analysis 40*(3), 493–517.

Barko, T., M. Cremers, and L. Renneboog (2020). Shareholder engagement on environmental, social, and governance performance. Working Paper.

Berg, F., J. Kölbel, and R. Rigobon (2020). Aggregate confusion: The divergence of ESG ratings. Working Paper.

Boffo, R. and R. Patalano (2020). ESG investing: Practices, progress and challenges. September 25, https://www.oecd.org/finance/ESG-Investing-Practices-Progress-Challenges.pdf.

Cao, J., H. Liang, and X. Zhan (2019). Peer effects of corporate social responsibility. *Management Science 65*(12), 5449–5956.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*(1), 57–82.

Cousins, P., M. Dutordoir, B. Lawson, and J. Q. F. Neto (2020). Shareholder wealth effects of modern slavery regulation. *Management Science*. Forthcoming.

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

Demers, E., J. Hendrikse, P. Joos, and B. I. Lev (2020). ESG didn't immunize stocks against the COVID-19 market crash. Working Paper.

Dimson, E., O. Karakaş, and X. Li (2020). Coordinated engagements. Working Paper.

Ding, W., R. E. Levine, C. Lin, and W. Xie (2020). Corporate immunity to the COVID-19 pandemic. *Journal of Financial Economics*. Forthcoming.

Eccles, R. G., I. Ioannou, and G. Serafeim (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science 60*(11), 2381–2617.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fauver, L., M. Hung, X. Li, and A. G. Taboada (2017). Board reforms and firm value: Worldwide evidence. *Journal of Financial Economics 125*(1), 120–142.

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

Flammer, C. (2020). Corporate green bonds. *Journal of Financial Economics*. Forthcoming.

Gibson, R., P. Krueger, and S. Mitali (2020). The sustainability footprint of institutional investors: ESG driven price pressure and performance. Working Paper.

Gibson, R., P. Krueger, and P. S. Schmidt (2019). ESG rating disagreement and stock returns. Working Paper.

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

Hassan, T. A., S. Hollander, L. van Lent, and A. Tahoun (2019). The global impact of Brexit uncertainty. Working Paper.

Krueger, P., D. Metzger, and J. Wu (2020). The sustainability wage gap. Working Paper.

Liang, H. and L. Renneboog (2017). On the foundations of corporate social responsibility. *Journal of Finance 72*(2), 853–909.

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

Ljungqvist, A., C. Malloy, and F. Marston (2009). Rewriting history. *Journal of Finance 64*(4), 1935–1960.

O'Donovan, J., H. F. Wagner, and S. Zeume (2019). The value of offshore secrets: Evidence from the Panama Papers. *Review of Financial Studies 32*(11), 4117–4155.

Refinitiv (2020a). Environmental, social and governance (ESG) scores from Refinitiv. April, https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/esg-scores-methodology.pdf.

Refinitiv (2020b). Refinitiv enhances ESG scoring methodology to reflect sustainable industry developments and market changes. April 15, https://www.refinitiv.com/en/media-center/press-releases/2020/april/refinitiv-enhances-esg-scoring-methodology-to-reflect-sustainable-industry-developments-and-market-changes.

Statman, M. and D. Glushkov (2009). The wages of social responsibility. *Financial Analyst Journal 65*(4), 33–46.

Tang, D. Y., J. Yan, and C. Y. Yao (2020). The determinants of ESG ratings: Rater ownership matters. Working Paper.

Ter Horst, J. and M. Verbeek (2007). Fund liquidation, self-selection, and look-ahead bias in the hedge fund industry. *Review of Finance 11*(4), 605–632.

Thomson Reuters (2017). ESG Thomson Reuters ESG Scores. March 29, https://web.archive.org/web/20170329145502/https://financial.thomsonreuters.com/content/dam/openweb/documents/pdf/financial/esg-scores-methodology.pdf.

WEF (2019). Seeking return on ESG: Advancing the reporting ecosystem to unlock impact for business and society (produced in collaboration with Allianz SE and Boston Consulting Group). January 24, http://www3.weforum.org/docs/WEF_ESG_Report_digital_pages.pdf.

Data Appendix

Table A1: Variable Definitions and Data Sources

This table defines the variables used in the analysis. CS-CIQ stands for "Compustat - Capital IQ" from Standard & Poor's.

Variable	Description	Source
ESG Variables		
ESG Score	Overall score of a firm's ESG performance. The score covers a firm's environmental (E), social (S) and corporate governance (G) performance. The score ranges between 0 (minimum score) and 100 (maximum score). The score is constructed using data $TRESGS$ in ASSET4 and data item $TR.TRESGScore$ in Refinitiv ESG.	Thomson Reuters Refinitiv ESG Data
E Score	Score of a firm's environmental performance. The score covers factors including a firm's resource use, emissions, and innovation. The score ranges between 0 (minimum score) and 100 (maximum score). The score is constructed using data item $ENVSCORE$ in ASSET4 and data item $TR.EnvironmentPillarScore$ in Refinitiv ESG.	Thomson Reuters Refinitiv ESG Data
S Score	Score of a firm's social performance. The score covers factors including workforce, human rights, community, and product responsibility. The score ranges between 0 (minimum score) and 100 (maximum score). The score is constructed using data item $SOCSCORE$ in ASSET4 and data item $TR.SocialPillarScore$ in Refinitiv ESG.	Thomson Reuters Refinitiv ESG Data
G Score	Score of a firm's corporate governance performance. The score covers factors including management, shareholders, and corporate social responsibility strategy. The score ranges between 0 (minimum score) and 100 (maximum score). The score is constructed using data item $CGVSCORE$ in ASSET4 and data item $TR.GovernancePillarScore$ in Refinitiv ESG.	Thomson Reuters Refinitiv ESG Data
Δ *ESG Score*	Percentage deviation in a firm's overall score between the initial (2018) and rewritten (2020) versions of the ESG data. The score deviation is computed for each firm-year combination as *ESG Score*$^{Rewritten\ Data}$ divided by *ESG Score*$^{Initial\ Data}$ minus one, times 100.	Thomson Reuters Refinitiv ESG Data
Δ *E Score*	Percentage deviation in a firm's E score between the initial (2018) and rewritten (2020) versions of the ESG data. The score deviation is computed for each firm-year combination as *E Score*$^{Rewritten\ Data}$ divided by *E Score*$^{Initial\ Data}$ minus one, times 100.	Thomson Reuters Refinitiv ESG Data

Variable	Description	Source
Δ *S Score*	Percentage deviation in a firm's S score between the initial (2018) and rewritten (2020) versions of the ESG data. The score deviation is computed for each firm-year combination as $S\ Score^{Rewritten\ Data}$ divided by $S\ Score^{Initial\ Data}$ minus one, times 100.	Thomson Reuters Refinitiv ESG Data
Δ *G Score*	Percentage deviation in a firm's G score between the initial (2018) and rewritten (2020) versions of the ESG data. The score deviation is computed for each firm-year combination as $G\ Score^{Rewritten\ Data}$ divided by $G\ Score^{Initial\ Data}$ minus one, times 100.	Thomson Reuters Refinitiv ESG Data
ES Treatment	Dummy variable that equals one for firms in the top quartile according to the E&S score in fiscal year 2017, and zero otherwise. The E&S score is calculated by averaging the *E Score* and the *S Score*.	Thomson Reuters Refinitiv ESG Data
Firm Characteristics		
Capex/Assets	Ratio of capital expenditures to total assets. The variable is constructed using Compustat data items $capx/at$. Winsorized at the 1% and 99% levels.	CS-CIQ, North America and Global, Fundamentals Annual
Cash/Assets	Ratio of cash plus short-term investments divided by total assets. The variable is constructed using Compustat data items csh/at. Winsorized at the 1% and 99% levels.	CS-CIQ, North America and Global, Fundamentals Annual
Debt/Assets	Ratio of total debt in current liabilities plus total long-term debt to total assets. The variable is constructed using Compustat data items $(dlc+dltt)/at$. Winsorized at the 1% and 99% levels.	CS-CIQ, North America and Global, Fundamentals Annual
EBIT/Assets	Ratio of earnings before interest and taxes to total assets. The variable is constructed using Compustat data items $ebit/at$. Winsorized at the 1% and 99% levels.	CS-CIQ, North America and Global, Fundamentals Annual
Log(Assets)	Logarithm of total assets. The variable is constructed using Compustat data item at. We use the U.S. Federal Reserve Board's H.10 Release to convert foreign currencies to USD. Winsorized at the 1% and 99% levels.	CS-CIQ, North America and Global, Fundamentals Annual, U.S. Federal Reserve Board's H.10 Release
PPE/Assets	Ratio of property, plant and equipment to total assets. The variable is constructed using Compustat data items $ppent/at$. Winsorized at the 1% and 99% levels.	CS-CIQ, North America and Global, Fundamentals Annual
R&D/Assets	Ratio of research and development expenses to total assets (missing values are set to zero). The variable is constructed using Compustat data items xrd/at. Winsorized at the 1% and 99% levels.	CS-CIQ, North America and Global, Fundamentals Annual
Sales Growth	Total sales at the end of the year divided by total sales at the end of the previous year, minus one. The variable is constructed using Compustat data items $[sale_t - sale_{t-1}]/sale_{t-1}$. Winsorized at the 1% and 99% levels.	CS-CIQ, North America and Global, Fundamentals Annual

Variable	Description	Source
Abnormal Return	Daily abnormal stock return calculated as the difference between the daily log-return of a stock (i.e., the logarithm of the gross return) and the CAPM beta times the daily log-return of the market. The CAPM beta is estimated using daily returns from 2017 and 2019. The market index is the S&P 500. Winsorized at the 1% and 99% levels.	CS-CIQ, North America
Post-COVID	Dummy variable that equals one from February 24 to March 31, 2020, and zero from January 1 to February 23, 2020.	Own calculations
Post-Fiscal	Dummy variable that equals one from March 18 to March 31, 2020, and zero from January 1 to 17 to March 17, 2020.	Own calculations

Figures

Figure 1: Cumulative Number of Academic Articles with ASSET4 Data Mentions over Time

This figure shows the cumulative number of academic articles (published papers and working papers) with ASSET4 data mentions over time. It also reports the cumulative number of articles that mention ASSET4 in combination with ESG. The data are retrieved from https://app.dimensions.ai/discover/publication on October 25, 2020. Our search terms are "ASSET4" and "ASSET4 AND ESG." The search is not case-sensitive. ASSET4 was founded in 2003. Most researchers continue to mention the term ASSET4 after the firm was acquired by Thomson Reuters in 2009.

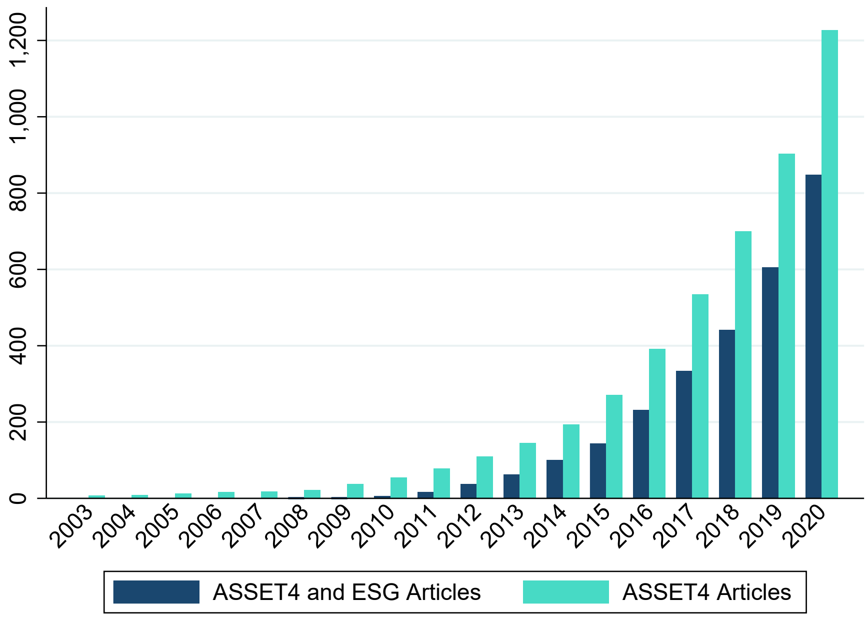

Figure 2: Distribution of Decile Rank Changes among S&P 1500 Firms

This figure shows the fractions of the sample that are subject to a decile rank change based on an ESG score in the 2018 (initial) and in the 2020 (rewritten) version of the data. The sample consists of firms in the S&P 1500 between 2011 and 2017. A value of 0 indicates no change in a firm's decile rank.

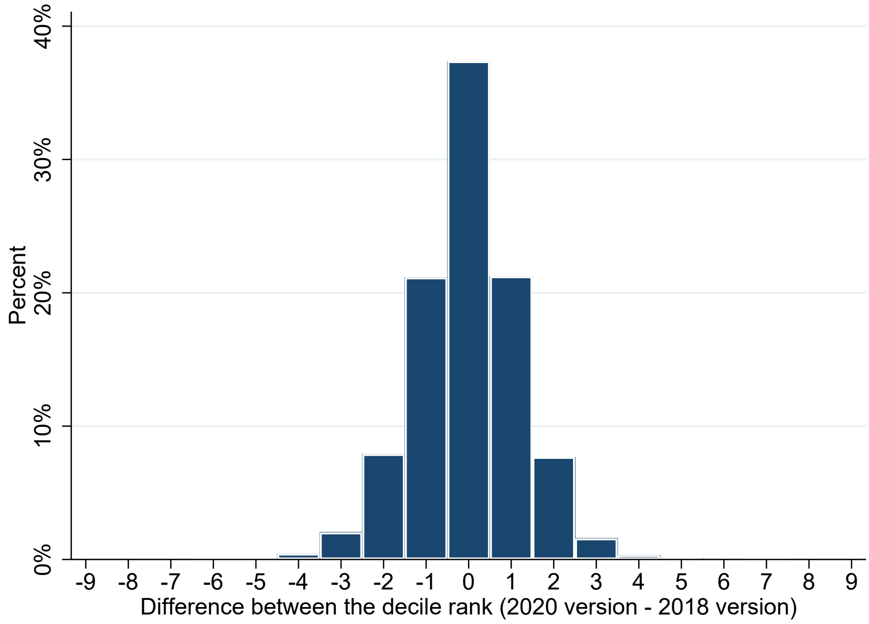

Figure 3: Decile Rank Changes among S&P 1500 Firms across Sample Years

This figure shows for each year the fractions of the sample that are subject to a decile rank change based on an ESG score in the 2018 (initial) and in the 2020 (rewritten) version of the data. The sample consists of firms in the S&P 1500 between 2011 and 2017. A value of 0 indicates no change in a firm's decile rank.

Tables

Table 1: Sample Composition over Time

This table reports information on the sample composition. The sample consists of 29,828 firm-year observations between 2011 and 2017 for which we have an ESG score in both the 2018 (initial) and the 2020 (rewritten) versions of the data. Column 1 reports the number of firm-year observations over time. Columns 2 and 3 report the number of firm-year observations for which an ESG score was added or deleted in the 2020 version.

Year	*ESG Score* in 2018 and 2020 Versions	*ESG Score* Additions in 2020 Version	*ESG Score* Deletions in 2020 Version
	(1)	(2)	(3)
2011	3,244	66	11
2012	3,386	69	11
2013	3,544	58	17
2014	3,737	49	14
2015	4,537	107	15
2016	5,418	154	16
2017	5,962	288	14
Full sample	29,828	791	98

Table 2: Deviations between the Initial and Rewritten Refinitiv ESG Data

This table documents deviations of the ESG scores between the 2018 (initial) and the 2020 (rewritten) versions of the Refinitiv ESG data. The sample consists of 29,828 firm-year observations between 2011 and 2017. Column 1 reports the fraction of firm-year observations that experienced an ESG score upgrade in the 2020 version (i.e., $\textit{ESG Score}^{\textit{Rewritten Data}} > \textit{ESG Score}^{\textit{Initial Data}}$). Column 2 reports the fraction of firm-year observations that experienced an ESG score downgrade in the 2020 version (i.e., $\textit{ESG Score}^{\textit{Rewritten Data}} < \textit{ESG Score}^{\textit{Initial Data}}$). $\textit{ESG Score}^{\textit{Rewritten Data}}$ is the ESG score in the 2020 data version and $\textit{ESG Score}^{\textit{Initial Data}}$ is the ESG Score in the 2018 data. The two columns add up to one since no single ESG score is the same across the two data versions. Columns 3 and 4 report mean and median values of a relative score deviation (in %), which is computed as $\textit{ESG Score}^{\textit{Rewritten Data}}$ divided by $\textit{ESG Score}^{\textit{Initial Data}}$ minus one, times 100. Columns 5 to 10 provide the same deviations for the environmental (E), social (S), and governance (G) subscores. Variable definitions are reported in Table A1.

	ESG Score		Δ *ESG Score*		Δ *E Score*		Δ *S Score*		Δ *G Score*	
Year	Up	Down	Mean	Median	Mean	Median	Mean	Median	Mean	Median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2011	12.1	87.9	-21.4	-19.0	-44.1	-39.7	-2.7	-19.0	113.6	-3.1
2012	13.6	86.4	-19.9	-17.5	-41.1	-36.2	0.2	-18.5	137.7	-4.7
2013	14.9	85.1	-20.1	-17.7	-42.1	-38.0	0.1	-19.0	131.3	-4.6
2014	15.1	84.9	-18.9	-16.7	-41.7	-38.2	4.6	-17.1	119.7	-4.9
2015	12.1	87.9	-20.7	-18.4	-48.3	-45.1	15.4	-17.0	110.4	-7.0
2016	11.4	88.6	-21.5	-19.6	-53.6	-52.1	14.7	-14.7	112.3	-9.4
2017	12.6	87.4	-20.9	-19.2	-53.0	-51.9	16.4	-12.0	102.4	-9.6
Full sample	13.0	87.0	-20.6	-18.4	-47.4	-44.0	8.6	-16.4	116.2	-6.9

Table 3: Summary Statistics

This table reports summary statistics for the sample. The sample consists of 29,828 firm-year observations common to both the initial and rewritten data between 2011 and 2017. Variable definitions are reported in Table A1.

Variable	Mean	STD	25%	Median	75%
$ESG\ Score^{Initial\ Data}$	50.41	18.04	36.13	49.90	64.51
$ESG\ Score^{Rewritten\ Data}$	41.82	20.64	25.23	39.94	57.54
$E\ Score^{Initial\ Data}$	50.35	31.94	16.04	48.24	84.45
$E\ Score^{Rewritten\ Data}$	32.78	28.86	4.10	27.66	56.78
$S\ Score^{Initial\ Data}$	50.52	31.43	18.73	50.02	82.51
$S\ Score^{Rewritten\ Data}$	42.12	23.57	23.44	39.57	59.72
$G\ Score^{Initial\ Data}$	50.09	30.24	21.21	53.26	77.50
$G\ Score^{Rewritten\ Data}$	48.46	22.75	30.08	48.76	66.82
Δ *ESG Score* (in %)	-20.57	21.26	-32.86	-18.43	-6.77
Δ *E Score* (in %)	-47.36	38.77	-80.97	-44.03	-19.69
Δ *S Score* (in %)	8.62	84.90	-37.29	-16.38	17.69
Δ *G Score* (in %)	116.24	353.88	-34.27	-6.95	84.36
Log(Assets)	8.54	1.77	7.42	8.47	9.61
Capex/Assets	0.05	0.05	0.02	0.03	0.06
Cash/Assets	0.15	0.15	0.04	0.10	0.19
Debt/Assets	0.24	0.18	0.10	0.23	0.36
EBIT/Assets	0.07	0.10	0.03	0.06	0.11
PPE/Assets	0.27	0.25	0.05	0.20	0.43
R&D/Assets	0.02	0.04	0.00	0.00	0.01
Sales growth	0.09	0.29	-0.02	0.05	0.14

Table 4: Variance Decomposition of the Deviations between the Initial and Rewritten Refinitiv ESG Data

This table reports a variance decomposition of the relative deviations (in %) in the total ESG score (Δ *ESG Score*) between the 2018 (initial) and the 2020 (rewritten) versions of the Refinitiv ESG data. The relative score deviation is computed for each firm-year combination as $ESG\ Score^{Rewritten\ Data}$ divided by $ESG\ Score^{Initial\ Data}$ minus one, times 100. We also provide the same analysis for the environmental (Δ *E Score*), social (Δ *S Score*), and governance (Δ *G Score*) subscores. The sample consists of 29,402 firm-year observations between 2011 and 2017 for which we have the industry and country data available. Regressions are estimated at the firm-year level. Industry fixed effects are based on the Fama-French 49 industry classification. Panel A shows the incremental R-squared obtained through the addition of the respective set of fixed effects to the specification. Panel B decomposes the variation termed "Firm level" from Panel A into firm fixed effects and the residual component. Variable definitions are reported in Table A1.

Panel A: Incremental R-Squared (in %)				
Dependent variable	Δ *ESG Score*	Δ *E Score*	Δ *S Score*	Δ *G Score*
Time Fixed Effects	0.2	1.8	0.8	0.1
Industry Fixed Effects	6.6	6.2	20.1	3.2
Industry x Time Fixed Effects	0.9	0.9	3.0	0.7
Country Fixed Effects	5.8	7.3	9.6	44.9
"Firm level"	86.5	83.8	66.5	51.1
Sum	100.0	100.0	100.0	100.0
Panel B: Fraction of Variation (in %)				
Dependent variable	Δ *ESG Score*	Δ *E Score*	Δ *S Score*	Δ *G Score*
Firm Fixed Effects	62.5	65.9	53.9	37.5
Residual	37.5	34.1	46.1	62.5
Sum	100.0	100.0	100.0	100.0

Table 5: Deviations between the Initial and Rewritten Refinitiv ESG Data and Firm Characteristics

This table reports the results of a regression of the relative deviations (in %) in the total ESG score ($\Delta\, ESGScore$) between the 2018 (initial) and the 2020 (rewritten) versions of the Refinitiv ESG data on a set of firm characteristics. The relative score deviation is computed for each firm-year combination as $ESG\ Score^{Rewritten\ Data}$ divided by $ESG\ Score^{Initial\ Data}$ minus one, times 100. We also report the same regressions for the environmental ($\Delta\, E\, Score$), social ($\Delta\, S\, Score$), and governance ($\Delta\, G\, Score$) subscores. The sample consists of 22,345 firm-year observations between 2011 and 2017 for which we have all data available. Regressions are estimated at the firm-year level. t-statistics, based on standard errors clustered at the firm level, are reported in parentheses. Variable definitions are reported in Table A1. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable	Δ *ESG Score*		Δ *E Score*	Δ *S Score*	Δ *G Score*
	(1)	(2)	(3)	(4)	(5)
Log(Assets)	4.80***	2.10***	9.16***	-7.75***	-17.92***
	(26.36)	(8.43)	(29.34)	(-12.93)	(-8.17)
Sales Growth	-5.30***	-3.54***	-7.09***	23.37***	20.64***
	(-10.31)	(-7.50)	(-7.60)	(9.62)	(4.40)
Capex/Assets	-10.91*	-12.78**	-10.81	12.09	-355.24***
	(-1.83)	(-2.31)	(-1.01)	(0.72)	(-5.15)
Cash/Assets	-0.81	-0.70	-8.65**	31.58***	126.09***
	(-0.41)	(-0.38)	(-2.26)	(4.24)	(3.91)
Debt/Assets	-2.38*	-0.54	-4.36	19.20***	-6.07
	(-1.65)	(-0.40)	(-1.64)	(3.52)	(-0.38)
EBIT/Assets	17.34***	10.31***	26.16***	-79.85***	69.04**
	(7.33)	(4.43)	(6.14)	(-8.78)	(2.15)
PPE/Assets	1.72	1.97	6.31**	-11.04**	50.98**
	(0.94)	(1.16)	(2.02)	(-1.99)	(2.52)
R&D/Assets	41.69***	20.46***	57.10***	134.14***	-572.91***
	(6.02)	(2.98)	(4.84)	(4.63)	(-5.62)
$ESG\ Score^{Initial\ Data}$		0.38***			
		(16.11)			
Observations	22,345	22,345	22,349	22,349	22,349
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes
Industry x year fixed effects	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.23	0.29	0.28	0.38	0.49

Table 6: Quantile Overlaps between the Initial and Rewritten ESG Refinitiv ESG Data

This table reports the fraction of firm-year observations (in %) assigned to different top and bottom quantiles (deciles, quartiles, and terciles) in both the 2018 (initial) and the 2020 (rewritten) versions of the Refinitiv ESG data. The sample consists of firms in the S&P 1500 between 2011 and 2017. We report quantile overlaps calculated based on 6,054 firm-year observations for which an ESG score (or its respective component part) is available in both versions of the data. Panel A shows the quantile overlaps for the total ESG score, while Panels B, C, and D show the quantile overlaps for the environmental (E), social (S), and governance (G) subscores, respectively. Panel E shows the quantile overlaps for an average of the environmental and social (E&S) subscores. Variable definitions are reported in Table A1.

	Panel A: *ESG Score*		
	Decile	Quartile	Tercile
Top	68.5	82.7	85.2
Bottom	62.4	76.8	80.5
	Panel B: *E Score*		
	Decile	Quartile	Tercile
Top	49.1	74.4	83.3
Bottom	92.7	89.9	88.1
	Panel C: *S Score*		
	Decile	Quartile	Tercile
Top	57.8	71.8	74.8
Bottom	37.7	60.7	67.7
	Panel D: *G Score*		
	Decile	Quartile	Tercile
Top	44.8	55.8	61.1
Bottom	42.6	54.0	59.9
	Panel E: *E&S Score*		
	Decile	Quartile	Tercile
Top	53.0	78.9	84.3
Bottom	45.6	73.3	80.0

Table 7: Returns during the Covid-19 Pandemic: Effects with Initial and Rewritten Refinitiv ESG Data

This table reports the results of difference-in-differences regressions for daily abnormal returns (*Abnormal Return*) during the first quarter of 2020 using either the 2018 (initial) or the 2020 (rewritten) versions of the Refinitiv ESG data. The sample consists of firms in the S&P 1500. *ES Treatment* equals one for firms ranked in the top quartile based on the E&S score in the year 2017, and zero otherwise. *ES Treatment* is constructed using either the 2018 (initial) or the 2020 (rewritten) versions of the Refinitiv ESG data. *Post-COVID* equals one from February 24, 2020 to March 31, 2020, and zero before this period. *Post-Fiscal* equals one from March 18, 2020 to March 31, 2020, and zero before this period. t-statistics, based on standard errors clustered by firm and day, are reported in parentheses. Variable definitions are reported in Table A1. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Baseline Results				
	2018 Data Version		2020 Data Version	
Dependent variable	*Abnormal Return*	*Abnormal Return*	*Abnormal Return*	*Abnormal Return*
	(1)	(2)	(3)	(4)
ES Treatment × *Post-COVID*	0.026	0.028	0.084**	0.086**
	(1.35)	(1.44)	(2.36)	(2.37)
ES Treatment	-0.014**		-0.024***	
	(-2.26)		(-3.33)	
Post-COVID	-0.300***		-0.316***	
	(-7.78)		(-9.00)	
Observations	81,163	81,162	81,163	81,162
Firm Fixed Effects	No	Yes	No	Yes
Day Fixed Effects	No	Yes	No	Yes
Adj. R-squared	0.017	0.022	0.017	0.022

Panel B: Controlling for Fiscal Stimulus				
	2018 Data Version		2020 Data Version	
Dependent variable	*Abnormal Return*	*Abnormal Return*	*Abnormal Return*	*Abnormal Return*
	(1)	(2)	(3)	(4)
ES Treatment × *Post-COVID*	0.007	0.009	0.096*	0.097*
	(0.31)	(0.39)	(1.73)	(1.74)
ES Treatment × *Post-Fiscal*	0.052**	0.052**	-0.030	-0.031
	(2.04)	(2.00)	(-0.51)	(-0.51)
ES Treatment	-0.014**		-0.024***	
	(-2.26)		(-3.33)	
Post-COVID	-0.189***		-0.212***	
	(-4.80)		(-6.29)	
Post-Fiscal	-0.303***		-0.283***	
	(-6.91)		(-7.47)	
Observations	81,163	81,162	81,163	81,162
Firm Fixed Effects	No	Yes	No	Yes
Day Fixed Effects	No	Yes	No	Yes
Adj. R-squared	0.024	0.022	0.024	0.022

Table 8: Announcement Returns around Refinitiv ESG's Methodology Change

This table reports cumulative abnormal returns (CARs) for S&P 1500 firms around Refinitiv ESG's methodology change announcement. The event date corresponds to March 6, 2020. CARs are measured over one-, two-, and three-day windows. We assign firms to two portfolios p, depending on the relative ESG score deviation for the year 2017 between the initial and revised data. The first portfolio is composed of firms in the top quartile of the relative score deviation ("Top Quartile"), and the second portfolio of firms in the bottom quartile ("Bottom Quartile"). The CAR for portfolio p is calculated as $T \times AR_p$, where T corresponds to the length of the event window. The return-generation process is based on the Fama-French-Carhart four-factor model. Column 1 (Column 2) tests whether CARs for firms in the top quartile (bottom quartile) are different from zero. Column 3 tests whether the CARs of a long-short portfolio of top minus bottom firms is different from zero. We estimate normal performance over an estimation window spanning 250 trading days before the event window. To be included in a portfolio, a firm must have a minimum of one hundred return observations in the estimation period and nonmissing return observations in all days of the event window. These conditions are met for 313 firms in the top-quartile portfolio and 298 in the bottom-quartile portfolio. We only consider firms for which the ESG score is available in both the initial and rewritten versions of the data. t-statistics, based on robust standard errors, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Event window	Top Quartile	Bottom Quartile	Top-Bottom Quartile
	(1)	(2)	(3)
[0;0]	0.0041***	-0.0052***	0.0093***
	(8.87)	(-10.94)	(18.68)
[-1;0]	0.0038	-0.0080***	0.0118**
	(1.13)	(-3.45)	(2.22)
[-1;1]	0.0070	-0.0167**	0.0237**
	(1.54)	(-2.53)	(2.55)

Internet Appendix

Table IA1: Studies using Refinitiv ESG (or ASSET4) Data

The table lists selected studies in leading finance journals as well as several recent working papers that use Refinitiv ESG (or ASSET4) data in their analyses.

Paper Authors	Paper Name	Paper Stage/Journal
Albuquerque et al. (2020)	Resiliency of environmental and social stocks: An analysis of the exogenous COVID-19 market crash	*Review of Corporate Finance Studies*
Bae et al. (2020)	Board reforms and dividend policy: International evidence	*Journal of Financial and Quantitative Analysis*
Cao et al. (2019)	Peer effects of corporate social responsibility	*Management Science*
Cousins et al. (2020)	Shareholder wealth effects of modern slavery regulation	*Management Science*
Dai et al. (2020)	Socially responsible corporate customers	*Journal of Financial Economics*
Ding et al. (2020)	Corporate immunity to the COVID-19 pandemic	*Journal of Financial Economics*
Eccles et al. (2014)	The impact of corporate sustainability on organizational processes and performance	*Management Science*
Fauver et al. (2017)	Board reforms and firm value: Worldwide evidence	*Journal of Financial Economics*
Ferrell et al. (2016)	Socially responsible firms	*Journal of Financial Economics*
Flammer (2020)	Corporate green bonds	*Journal of Financial Economics*
Liang and Renneboog (2017)	On the foundations of corporate social responsibility	*Journal of Finance*
O'Donovan et al. (2019)	The value of offshore secrets: Evidence from the Panama Papers	*Review of Financial Studies*
Barko et al. (2020)	Shareholder engagement on environmental, social, and governance performance	Working Paper
Berg et al. (2020)	Aggregate confusion: The divergence of ESG ratings	Working Paper
Dimson et al. (2020)	Coordinated engagements	Working Paper
Demers et al. (2020)	ESG didn't immunize stocks against the COVID-19 market crash	Working Paper
Gibson et al. (2020)	The sustainability footprint of institutional investors: ESG driven price pressure and performance	Working Paper
Krueger et al. (2020)	The sustainability wage gap	Working Paper

Table IA2: Absolute ESG Scores Deviations between the Initial and Rewritten Data Versions

This table documents absolute deviations of the total ESG score between the 2018 (initial) and the 2020 (rewritten) versions of the ESG data. We also provide the same deviations for the environmental (E), social (S), and governance (G) subscores. The sample consists of 29,828 firm-year observations between 2011 and 2017. The absolute score deviation is computed for each firm-year combination as $ESG\ Score^{Rewritten\ Data}$ minus $ESG\ Score^{Initial\ Data}$. All ESG data items are expressed in points. Variable definitions are reported in Table A1.

	Δ *ESG Score*		Δ *E Score*		Δ *S Score*		Δ *G Score*	
Year	Mean	Median	Mean	Median	Mean	Median	Mean	Median
2011	-8.8	-9.1	-16.4	-12.5	-10.3	-8.4	-0.7	-1.8
2012	-8.1	-8.4	-15.5	-11.6	-9.8	-8.5	-1.2	-2.8
2013	-8.2	-8.3	-15.9	-11.9	-9.7	-8.2	-0.9	-2.9
2014	-7.9	-8.1	-15.9	-12.1	-9.0	-7.8	-0.5	-3.1
2015	-8.7	-8.9	-18.0	-13.7	-8.1	-8.1	-1.2	-4.3
2016	-9.1	-9.3	-19.7	-15.1	-7.5	-7.4	-2.7	-5.5
2017	-8.9	-9.1	-19.2	-14.6	-6.4	-6.1	-2.9	-5.7
Full sample	-8.6	-8.8	-17.6	-13.8	-8.4	-7.6	-1.6	-4.2

Table IA3: Description of Changes to the ESG Scoring Methodology

The table cites the description of the changes to Refinitiv ESG's scoring methodology (Refinitiv, 2020b).

Change name	Description provided by Refinitiv
(1) Change to Materiality Matrix	"Refinitiv enhanced ESG scores further takes into account that not all metrics have the same importance to every industry. The Refinitiv ESG magnitude matrix is developed as a proprietary model and is applied at the category level. Importantly, the magnitude values are automatically and dynamically adjusted as ESG corporate disclosure evolves and matures. For Boolean metrics, levels of data disclosure can act as a proxy for investor driven pressure on company reporting. Levels of disclosure inform the relative 'weight' of data points for each industry. For measurable numeric metrics, we use our data to determine which sectors contribute most and the proportion of the contribution to the total is used as a proxy for the level of materiality for that sector. For example, the more a given sector contributes to carbon emissions, the more material are carbon emissions data points to companies in that sector. Refinitiv proprietary "magnitude matrix" assesses materiality, showing the weight, from 1 to 10, of data points for each industry."
(2) Change to Transparency/ Investment Grade Scores	"The previous ESG scoring methodology allocated a score of 0.5 to companies which didn't report on metrics, essentially giving them the 'benefit of the doubt'. However, as this may disincentivize companies to report on their ESG performance, the enhanced methodology assigns a score of zero to companies who don't report on metrics relevant to the industry. This new approach encourages company disclosure and transparency."

Table IA4: Country Distribution of the Sample

This table lists the country breakdown of the sample firms for which an ESG score is available in the initial or the rewritten data versions. The country code corresponds to data item *loc*, which indicates the country of headquarters in the Standard & Poor's Compustat - Capital IQ database.

Country Code	Freq.	Percent	Country code	Freq.	Percent
ARE	73	0.2	KOR	760	2.5
ARG	63	0.2	KWT	49	0.2
AUS	2,010	6.5	LKA	7	0.0
AUT	113	0.4	LUX	67	0.2
BEL	179	0.6	MAC	21	0.1
BHR	21	0.1	MAR	14	0.1
BMU	196	0.6	MCO	4	0.0
BRA	528	1.7	MEX	224	0.7
CAN	1,751	5.7	MYS	356	1.2
CHE	480	1.6	NGA	7	0.0
CHL	183	0.6	NLD	247	0.8
CHN	974	3.2	NOR	125	0.4
COL	94	0.3	NZL	204	0.7
CYM	28	0.1	OMN	33	0.1
CYP	7	0.0	PAK	5	0.0
CZE	23	0.1	PAN	9	0.0
DEU	600	2.0	PER	55	0.2
DNK	176	0.6	PHL	168	0.6
EGY	63	0.2	PNG	14	0.1
ESP	305	1.0	POL	197	0.6
FIN	160	0.5	PRT	59	0.2
FRA	625	2.0	QAT	61	0.2
GBR	2,058	6.7	ROU	2	0.0
GEO	2	0.0	RUS	228	0.7
GRC	121	0.4	SAU	66	0.2
HKG	808	2.6	SGP	283	0.9
HUN	28	0.1	SVN	1	0.0
IDN	239	0.8	SWE	380	1.2
IND	616	2.0	THA	221	0.7
IRL	138	0.5	TUR	182	0.6
ISR	109	0.4	TWN	909	3.0
ITA	312	1.0	URY	2	0.0
JEY	7	0.0	USA	9,056	29.5
JOR	7	0.0	VGB	1	0.0
JPN	2,799	9.1	ZAF	794	2.6
KEN	3	0.0	ZWE	7	0.0

Table IA5: Industry Distribution of the Sample

This table lists the industry breakdown of the sample firms for which with an ESG score is available in the initial or rewritten data versions. We use the Fama-French 49 industry classification derived from the data item *SIC*, which indicates the Standard Industrial Classification (SIC) code in the Standard & Poor's Compustat - Capital IQ database. Firm-year observations (a total of 431) with the following SIC codes or a missing SIC code are not contained in any category: 900, 3990, 6797, 9995, 9997, 9998.

Industry	Freq.	Percent	Industry	Freq.	Percent
1 Agriculture	77	0.25	26 Defense	28	0.09
2 Food Products	740	2.44	27 Precious Metals	500	1.65
3 Candy and Soda	101	0.33	28 Mining	729	2.41
4 Beer and Liquor	227	0.75	29 Coal	177	0.58
5 Tobacco Products	74	0.24	30 Petroleum and Natural Gas	1,373	4.53
6 Recreation	97	0.32	31 Utilities	1,306	4.31
7 Entertainment	321	1.06	32 Communication	1,168	3.86
8 Printing and Publishing	199	0.66	33 Personal Services	209	0.69
9 Consumer Goods	450	1.49	34 Business Services	1,015	3.35
10 Apparel	298	0.98	35 Computers	331	1.09
11 Healthcare	274	0.90	36 Computer Software	1,270	4.19
12 Medical Equipment	376	1.24	37 Electronic Equipment	1,019	3.36
13 Pharmaceutical Products	1,129	3.73	38 Measuring and Control Equip.	260	0.86
14 Chemicals	864	2.85	39 Business Supplies	272	0.90
15 Rubber and Plastic Products	104	0.34	40 Shipping Containers	111	0.37
16 Textiles	61	0.20	41 Transportation	1,232	4.07
17 Construction Materials	644	2.13	42 Wholesale	743	2.45
18 Construction	742	2.45	43 Retail	1,474	4.87
19 Steel Works	600	1.98	44 Restaurants, Hotels, Motels	347	1.15
20 Fabricated Products	36	0.12	45 Banking	2,552	8.43
21 Machinery	829	2.74	46 Insurance	1,097	3.62
22 Electrical Equipment	278	0.92	47 Real Estate	841	2.78
23 Automobiles and Trucks	652	2.15	48 Trading	2,502	8.26
24 Aircraft	171	0.56	49 Other	251	0.83
25 Shipbuilding, Railroad Equip.	135	0.45			

Table IA6: Quantile Overlaps between the Initial and Rewritten Data Versions for S&P 1500 Firms

This table reports the fraction of firm-year observations (in %) assigned to different top and bottom quantiles (deciles, quartiles, and terciles) in either the 2018 (initial) and the 2020 (rewritten) versions of the Refinitiv ESG data. The sample includes firms in the S&P 1500 firms between 2011 and 2017. We report the quantile overlaps calculated based on 6,524 firm-year observations for which an ESG score (or its respective component part) is available in either version of the data. Panel A shows the quantile overlaps for the total ESG score, while Panels B, C, and D show the quantile overlaps for the environmental (E), social (S), and governance (G) subscores, respectively. Panel E shows the quantile overlaps for an average of the environmental and social (E&S) subscores. Variable definitions are reported in Table A1.

Panel A: *ESG Score*			
	Decile	Quartile	Tercile
Top	63.3	76.2	78.8
Bottom	58.1	71.2	75.0
Panel B: *E Score*			
	Decile	Quartile	Tercile
Top	28.4	52.3	59.5
Bottom	75.7	74.9	75.0
Panel C: *S Score*			
	Decile	Quartile	Tercile
Top	53.0	66.3	69.2
Bottom	35.6	56.7	63.0
Panel D: *G Score*			
	Decile	Quartile	Tercile
Top	42.0	52.2	57.0
Bottom	38.8	50.4	55.6
Panel E: *E&S Score*			
	Decile	Quartile	Tercile
Top	48.1	72.7	77.7
Bottom	43.1	68.2	74.3

about ECGI

The European Corporate Governance Institute has been established to improve *corporate governance through fostering independent scientific research and related activities.*

The ECGI will produce and disseminate high quality research while remaining close to the concerns and interests of corporate, financial and public policy makers. It will draw on the expertise of scholars from numerous countries and bring together a critical mass of expertise and interest to bear on this important subject.

The views expressed in this working paper are those of the authors, not those of the ECGI or its members.

www.ecgi.global

Electronic Access to the Working Paper Series

The full set of ECGI working papers can be accessed through the Institute's Web-site (www.ecgi.global/content/working-papers) or SSRN:

Finance Paper Series	http://www.ssrn.com/link/ECGI-Fin.html
Law Paper Series	http://www.ssrn.com/link/ECGI-Law.html