Drawing Up the Bill: Are ESG Ratings Related to Stock Returns Around the World?

Rómulo Alves[†] Philipp Krueger[‡] Mathijs van Dijk[§]

January 13, 2025

Abstract

We provide the most comprehensive analysis to date of the relation between ESG ratings and stock returns, using 16,000+ stocks in 48 countries and seven different ESG rating providers. We find very little evidence that ESG ratings are related to global stock returns between 2001 and 2020. This finding obtains across different regions, time periods, ESG (sub)ratings, ESG momentum, ESG downgrades and upgrades, and best-in-class strategies. We further find little empirical support for prominent hypotheses from the literature on the role of ESG uncertainty and of country-level ESG social norms, ESG disclosure standards, and ESG regulations in shaping the relation between ESG and global stock returns. Overall, our results suggest that ESG investing did not systematically affect investment performance during the past two decades.

JEL Classifications: G11; G12; G15

Keywords: ESG investing; Environmental, Social & Governance; Global stock returns; ESG uncertainty; Country characteristics

[*]We thank Matthew Ringgenberg (the editor), two anonymous referees, Veronika Czellar, Tom Gosling, Daniel Metzger, Loriana Pelizzon, Raghavendra Rau, Luc Renneboog, Henri Servaes, Marta Szymanowska, David Thesmar, and seminar participants at the Dutch Banking Association, the Erasmus Initiative for Dynamics of Inclusive Prosperity, Jakarta State University, Manchester Business School, PGGM, Robeco Asset Management, Tilburg University, and the 2021 Netspar annual conference for helpful comments and suggestions. We are grateful to the Network for Studies on Pensions, Aging and Retirement (Netspar) and to the Erasmus Trustfonds / Stichting Tot Organisatie van Effectenhandelaren Rotterdam (STOER) for financial support.

[†]SKEMA Business School - Université Côte d'Azur. E-mail: romulo.alves@skema.edu

[‡]University of Geneva, Swiss Finance Institute & European Corporate Governance Institute. E-mail: philipp.krueger@unige.ch

[§]Rotterdam School of Management, Erasmus University & Netspar. E-mail: madijk@rsm.nl.

1 Introduction

Recent decades have witnessed a remarkable growth in "Environmental, Social, and Governance" (ESG) investing. According to GSIA (2021), global ESG investment reached US$35.3 trillion (over one third of global assets under management) in 2020. The question of whether incorporating ESG considerations into investment strategies helps or hurts financial performance has occupied the minds of academics and investors alike.

A popular view in both academic research and the financial industry is that investors can "do well while doing good." In a survey of over 2,000 empirical studies, Friede, Busch, and Bassen (2015) conclude that "the business case for ESG investing is empirically very well founded." Indeed, improving financial returns is a key motivation for ESG investing among financial institutions (Amel-Zadeh and Serafeim 2018, Dyck, Lins, Roth, and Wagner 2019, BNP Paribas 2021).

Yet, there are reasons to be skeptical about a consistently positive relation between ESG and stock returns. First, many underlying studies are limited in scope. They use ESG ratings from a single provider, even though there can be considerable differences across providers (Brandon, Krueger, and Schmidt 2021, Berg, Koelbel, and Rigobon 2022). They use short sample periods, from specific markets, even though results may then be driven by temporary investor and consumer demand effects (Pástor, Stambaugh, and Taylor 2022, Van der Beck 2024). They use data on U.S. stocks only even though the majority of global sustainable assets are elsewhere (GSIA 2021). Second, several prominent papers indicate that sustainability may be inversely related to stock returns (Hong and Kacperczyk 2009, Chava 2014, Bolton and Kacperczyk 2021, 2023), in line with theoretical predictions (Fitzgibbons, Pedersen, and Pomorski 2021, Pástor, Stambaugh, and Taylor 2021). In a recent literature review, Liang and Renneboog (2021) thus conclude that "there is still no consensus about [whether] ESG-based investing helps or hurt performance."

We aim to synthesize the evidence by using a comprehensive database covering 16,368 unique stocks traded in 48 countries between 2001 and 2020, and by using ESG ratings from seven major rating agencies. To our knowledge, this is the most comprehensive database assembled to date to study the relation between ESG ratings and stock returns.[1]

[1] After posting a first version of this paper in June 2022, we became aware of several related papers, including Berg, Lo, Rigobon, Singh, and Zhang (2023), Karolyi, Wu, and Xiong (2023), Lindsey, Pruitt, and Schiller (2023), and Eskildsen, Ibert, Jensen, and Pedersen (2024). Please see below for a discussion of how we deviate from these papers.

We start out by examining whether ESG ratings predict cross-sectional variation in stock returns in our global sample. We do so by estimating monthly Fama and MacBeth (1973) cross-sectional regressions of the returns on individual stocks on each of the seven individual ESG ratings (FTSE, ISS, MSCI IVA, Refinitiv, RepRisk, S&P Global, and Sustainalytics) as well as on two versions of a "composite" ESG measure that is computed as the average ESG by stock across multiple ESG ratings. In line with Berg et al. (2022), we find that correlations across ESG ratings of different raters are also low globally, at around 0.5 to 0.6. Our regressions include a considerably more extensive set of stock-level control variables (market beta, size, book-to-market, investment, profitability, momentum, volatility, leverage, tangibility, R&D) than in many other studies, to preclude that any ESG effect is driven by one of the other well-known cross-sectional return predictors.

Our main finding is that there is very little evidence that ESG ratings are related to future stock returns (controlling for other stock characteristics). This finding holds across the seven individual ESG ratings and the two composite ESG ratings. Since there are strong country- and industry-level components in ESG ratings (Gillan, Koch, and Starks 2021), we rerun our analyses using country- and/or industry-adjusted ESG ratings and obtain similar results.

We further estimate our regressions separately for different geographic regions; for different subperiods of our full sample period 2001-2020; for the E, S, and G subratings individually; for ESG momentum (Bekaert, Rothenberg, and Noguer 2023); for ESG upgrades and downgrades (Krueger 2015, Shanaev and Ghimire 2022); and also for best-in-class ESG strategies (Statman and Glushkov 2009) and find no consistent evidence of a relation between ESG ratings and stock returns.

The pervasive failure to reject the null hypothesis of no relation between ESG ratings and stock returns in our battery of tests could stem from a genuinely weak relation in our sample, but also from insufficient statistical power to identify potentially sizable effects. For this reason, we carry out an extensive analysis of the effect size (economic significance) of the estimated coefficients on the ESG ratings as well as of the statistical power of our tests. Overall, coefficient point estimates are economically small; they generally indicate that a one standard deviation shock to ESG ratings in our sample is associated with a change in returns of less than 1% per annum – and often considerably smaller. Our statistical power analysis indicates that, although we are unable to state that power is never a concern (especially for tests based on subsamples of our overall sample), a lack of statistical power is unlikely to explain the general absence of a significant relation between ESG ratings and future stock returns in the data. Thus, our overall failure to reject the null of no relation between ESG ratings and stock returns is likely due to the relation truly being weak.

Since Fama-MacBeth regressions take an investor perspective in the sense that the ESG rating coefficients in the monthly cross-sectional regressions can be interpreted as the monthly returns on a zero-investment portfolio that is invested according to the ESG rating signal (and since we take appropriate lags into account such that the strategy is implementable in practice; Zhang 2023), our evidence thus suggests that incorporating ESG ratings into investment strategies did not systematically affect investment performance over 2001-2020. We obtain similar results when we estimate value-weighted Fama-MacBeth regressions (to preclude that the results are driven by microcaps) and when we use portfolio sorts instead of Fama-MacBeth regressions.

One potential reason for the lack of a relation between ESG ratings and stock returns is that investors may be uncertain about the "true" ESG rating of a stock, given the difficulties in measuring a firm's ESG performance (Berg et al. 2022). Avramov, Cheng, Lioui, and Tarelli (2022) present a model in which ESG uncertainty weakens the relation between ESG ratings and stock returns, and find support for the model's predictions in U.S. data. Using our global sample with seven different ESG ratings, we re-estimate their regressions of stock returns on ESG and an interaction term of ESG with an indicator variable for stocks with low ESG uncertainty. In line with their empirical findings and their model, we obtain a significantly negative coefficient on the interaction term for U.S. stocks in the first half of our sample period. However, the interaction coefficient is not significant for the U.S. in the second half of our sample period (consistent with Avramov et al. 2022) nor for any other major geographic region in either the first or the second half of our sample period. In sum, this analysis suggests that ESG uncertainty can shed only limited light on the lack of a relation between ESG ratings and stock returns around the world.

The literature has also put forward various hypotheses on country characteristics that could moderate the relation between ESG ratings and stock returns. First, in countries with social norms reflecting more positive attitudes and beliefs regarding ESG issues, investors may have stronger ESG preferences which may then be more likely to be priced in the stock market (Dyck et al. 2019). Second, the incorporation of ESG information may be hampered by the poor quality of ESG disclosure standards (Ilhan, Krueger, Sautner, and Starks 2023), which vary considerably across countries (Krueger, Sautner, Tang, and Zhong 2024). Third, in countries with stricter ESG regulations, investors may be more concerned about the potential financial consequences of ESG-related risks (Bolton and Kacperczyk 2023). To assess these three hypotheses, we follow recent studies and collect country-level data on a considerable number of proxies for ESG social norms, ESG disclosure standards, and the strictness of ESG regulations. We then estimate regressions of global stock returns on ESG ratings while adding interaction terms of the ESG ratings with these various country characteristics. We find little evidence that these hypotheses are helpful in understanding cross-country variation in the relation between ESG ratings and returns.

Overall, our results indicate that there is no evidence of a statistically or economically significant relation between ESG ratings and global stock returns over 2001-2020. This finding could be viewed as comforting by investors as it suggests that incorporating ESG ratings into investment strategies has not come at the expense of financial returns in the past two decades. It also suggests that the risk of "green bubbles" due to large ESG investment flows may be limited. That said, our analysis suggests that ESG investing on the basis of ESG ratings has so far not been effective in reducing (increasing) the cost of equity capital of strong (poor) ESG firms (Berk and Van Binsbergen 2025), which could lead firms to internalize climate and social externalities (Fama 2021, Pástor et al. 2021).

Our main contribution is to synthesize the evidence on the relation between ESG ratings and stock returns using a large global database, multiple ESG ratings, and appropriate statistical methods, thereby showing that prior results suggesting a link between ESG and stock returns do not systematically hold in a global sample. We do not mean to imply that either the academic community or the financial industry holds strong and consistent views that directly oppose our main finding. We also do not aim to refute specific individual studies and/or explain why prior studies obtain different results – which seems infeasible given the vast literature on this question. However, given the large heterogeneity in the results reported in this literature, we believe that our comprehensive analysis highlighting the lack of a relation between ESG ratings and returns may be of interest to academics and practitioners alike. Further, our study does point out that two prominent results in the literature (the moderating effect of ESG rating uncertainty on the ESG-return relation documented by Avramov et al. (2022) and the "greenium" documented by Pástor et al. (2022)) do not obtain for countries other than the United States, other time periods, and/or other ESG ratings.

Our analysis does not rule out the possibility that there may be more subtle empirical patterns that we do not uncover. Indeed, there may be a significant relation with stock returns for more specific variables and subcomponents of ESG pillars, specific time periods, specific types of firms, specific countries, and/or under specific conditions. Yet, in our view, the consistency of our evidence of a lacking relation between ESG ratings and stock returns across different rating agencies, regions, time periods, ESG (sub)ratings, ESG momentum, ESG downgrades and upgrades, and best-in-class strategies does raise the bar for the evidence on such specific findings, in particular in light of the multiple hypothesis testing problem (Harvey, Liu, and Zhu 2016, Heath, Ringgenberg, Samadi, and Werner 2023b).

As per the return decomposition of Campbell (1991), realized stock returns are the sum of expected returns, cash flow shocks, and discount rate shocks. Our study of the relation between realized stock returns and ESG ratings thus does not directly speak to the question of whether ESG ratings might be related to expected returns or to cash flow and/or discount rate shocks. This question is important, especially in light of the challenge of disentangling short-term repricing dynamics (for example, because of temporary investor demand effects) from long-term expected return effects. We consider addressing this question to be out of the scope of the current paper (in part also because estimating expected returns and cash flow shocks is fraught with a host of empirical challenges), but such an analysis would be a fruitful avenue for further research.

Our focus on ESG ratings is motivated by the central role they play in the financial industry and in particular as drivers of investment flows (Amel-Zadeh and Serafeim 2018, Hartzmark and Sussman 2019, Dell'Erba and Doronzo 2023).[2] Our evidence speaks to the ongoing debate on whether ESG ratings have material effects on stock returns and whether they lead financial markets to allocate capital to more sustainable firms (Heath, Macciocchi, Michaely, and Ringgenberg 2023a). This debate is particularly relevant in light of recent policy discussions on the quality of ESG ratings (IOSCO 2021, European Commission 2023).

We deviate from related efforts by Berg et al. (2023), Karolyi et al. (2023), Lindsey et al. (2023), and Eskildsen et al. (2024) as – to the best of our knowledge – we are the only paper that combines a large global sample of stocks with a substantial number of different ESG ratings and an extensive set of stock-level control variables, while also examining ESG momentum, ESG upgrades and downgrades, best-in-class strategies as well as several hypotheses on how ESG uncertainty

[2]Other studies focus on the relation between stock returns and more specific variables such as carbon emissions (Bolton and Kacperczyk 2021, 2023, Zhang 2023) and employee satisfaction (Edmans, Pu, Zhang, and Li 2024b).

and country characteristics could affect the relation between ESG and returns. Berg et al. (2023) show that combining ratings from six ESG rating agencies using inventive statistical and voting aggregation techniques can produce portfolios with positive alphas in the United States, Europe and Japan from 2014 to 2020. Although we acknowledge that it may be possible to attenuate the noise in ESG ratings using these and other approaches, our main interest is in the "plain" relation between ESG ratings and stock returns – since we believe these aproaches are currently not widely used in practice. Consistent with our findings, Karolyi et al. (2023) conclude that the greenium is largely limited to North America. We add to their paper by also examining the relation between the S and G subratings as well as the overall ESG rating with global stock returns, by using multiple rating agencies, and by testing a number of additional hypotheses on whether and how ESG ratings could be linked to stock returns. Lindsey et al. (2023) use data from seven ESG rating agencies to show that instrumented principal components analysis can be used to generate ESG investment strategies that do not cost any financial return – in line with our finding of an insignificant ESG-return relation. Their paper is limited to the U.S. and does not examine the additional hypotheses we test in our paper. Eskildsen et al. (2024) focus only on the U.S. and the environmental dimension. Consistent with our evidence, they estimate an insignificant greenium when using realized returns and accounting for multiple hypothesis testing.

2 Data

We construct a global database of monthly stock returns and characteristics covering the period from January 2001 to December 2020 from the Center for Research in Security Prices (CRSP), Compustat North America, and Compustat Global. These databases are survivorship bias-free and jointly cover over 98% of worldwide market capitalization. We clean the data following Bessembinder, Chen, Choi, and Wei (2019) and Chaieb, Langlois, and Scaillet (2021). Our analyses include the following stock-level control variables: market beta, size, book-to-market, investment, profitability, momentum, volatility, leverage, tangibility, and R&D (Hou, Kho, and Karolyi 2011, Fama and French 2015, Bolton and Kacperczyk 2021). We match stock return data between July of year *t+1* and June of year *t+2* to accounting data available at the end of year *t*. We winsorize all control variables at the 0.5% and 99.5% levels based on the whole sample distribution. We refer to the

Internet Appendix for a detailed explanation of the data filters, variable definitions, and sources as well as for detailed summary statistics.

We use the following seven different ESG ratings: FTSE, ISS, MSCI Intangible Value Assessment (MSCI IVA), Refinitiv, RepRisk, S&P Global, and Sustainalytics – which have broad global coverage and are among the most widely used by investors (SustainAbility 2020). We re-scale each ESG rating to range from zero to 100, where 100 indicates the best ESG performance. We use both the RepRisk index (RRI) and the RepRisk rating (RRR), which, in contrast to the other ratings, quantify firms' reputational risk exposure to ESG issues – and thus we invert both ratings to increase with ESG performance (decrease with ESG reputational risk).[3] We also use the separate environmental (E), social (S), and governance (G) subratings from each rater, re-scaled to range from zero to 100. We match stock return data in a particular month with the most recent ESG ratings available in the previous month to avoid look-ahead bias (Zhang 2023). For raters that provide annual instead of monthly ESG ratings, we assume that investors observe the rating with a delay of six months.

To examine whether aggregating ESG ratings across different raters could potentially reduce noise and strengthen the relation with stock returns, we construct two versions of a "composite" ESG measure that is the average by stock across multiple ESG ratings. We do not include RepRisk in these composite ratings since RepRisk ratings differ both conceptually and empirically from the other ratings, as discussed below (but our findings do not change when including RepRisk). *Composite 6* is the average rating across all six raters (other than RepRisk) and is missing for stocks not covered by all six raters raters in a specific month. Since *Composite 6* is characterized by a lot of missing values, we also compute *Composite 3+* as the average rating across at least three raters (other than RepRisk) and is missing for stocks covered by fewer than three raters in a specific month. Because the statistical distributions differ across raters, we follow Gillan et al. (2021) and convert the ratings at each point in time to percentile ranks before averaging. We note that the coverage differs substantially across raters and, as a result, we impose a starting date of January 2014 for the composite ratings.

Our final final database covers 16,368 stocks traded in 48 countries from January 2001 to December 2020. We impose the starting date for each rating to be the first date for which at least 120

[3]RRR adjusts the RRI by taking into account both the sector in which the firm operates and the countries where firms are headquartered and experienced ESG risk incidents. We convert the ten RRR letter rankings to a range from zero to 100 such that 100 corresponds to the lowest ESG reputational risk.

stocks are rated. When we conduct analyses by geographic region, we also impose this restriction to ensure a reasonable minimum number of observations.

Figure 1 shows how the stocks in our sample are distributed across geographic regions and countries. 68% of the stocks in our sample are traded in developed countries and 32% in emerging countries. North American stock exchanges are home to 34% of sample stocks, followed by the regions Emerging Countries (32%) and Europe (16%). Japan and Asia-Pacific each account for 9% of the sample. The countries that represent the greatest number of sample stocks are the U.S. (30%), China (10%), Japan (9%), South Korea (5%), United Kingdom (5%), India (5%), Canada (4%), Australia (4%), Hong Kong (4%), and Taiwan (3%).

Figure 2 shows pooled correlations across the different ESG ratings. In particular, this figure shows the average correlations across all rater-pairs for five major geographic regions (Asia-Pacific, Emerging Countries, Europe, Japan, North America) and for the global sample, as well as the average correlation excluding RepRisk, and the minimum (and minimum excluding RepRisk) and maximum correlation across rater-pairs by region. The Internet Appendix presents the same correlations separately for the E, S, and G subratings. Three findings emerge. First, both globally and for each region, the average correlations across ESG ratings are far from perfect; they range from around 0.5 to 0.6 (0.3 to 0.4) if we exclude (include) RepRisk – extending the findings of Berg et al. (2022) to a global sample. In line with Berg et al. (2023), we find that, remarkably, the RepRisk ratings are negatively correlated with the other ratings even while they have an inverted scale. Second, there is substantial variation in correlations across rater pairs, with global ESG rating correlations reaching a maximum of 0.73 and a minimum of 0.37 (-0.40) if we exclude (include) RepRisk. Third, average correlations across raters (excluding RepRisk) are highest for the overall *ESG* rating (0.59), followed by the *E* (0.56), *S* (0.49), and *G* (0.33) subratings.[4]

[4]This pattern may be expected since *ESG* ratings likely average out noise and *E* tends to be more objectively measured than *S* and *G* (Gillan et al. 2021). For example, whereas carbon emissions are an objective measure raters agree on, it is less clear whether gender parity is better captured by gender seniority gaps, gender pay gaps, or the number of sexual harassment lawsuits. *G* is also prone to disagreement given the lack of agreement on what constitutes good corporate governance and the possibility that optimal governance arrangements may vary across countries (Black, Carvalho, and Érica Gorga 2012).

3 Results

In this section, we examine the relation between ESG ratings and stock returns (Section 3.1), we analyze effect sizes and the statistical power of our tests (Section 3.2), we study whether ESG uncertainty weakens the relation between ESG ratings and stock returns (Section 3.3), we assess three hypotheses from the literature on country characteristics that could moderate the relation between ESG ratings and stock returns (Section 3.4), and we consider the issue of measurement error (Section 3.5). In Section 3.6, we discuss our results in light of the existing literature.

3.1 Are ESG ratings related to stock returns?

Table 1 presents the results of monthly Fama-Macbeth cross-sectional regressions of stock returns on lagged ESG ratings and control variables. Each column (1) through (10) shows the average coefficient (and associated Newey and West (1987) t-statistic with automatic lag selection in parentheses) for the eight individual ESG ratings (including *RRI* and *RRR* from RepRisk) and the *Composite 3+* and *Composite 6* ratings, as well as for all control variables in the model. The bottom five rows show the 95% confidence interval (95% CI) of the coefficient point estimates corresponding to each ESG rating, the number of stock-month observations included in the regressions, the average R^2, and the first month in the sample for which that particular rating is available.

The main finding in Table 1 is that there is very little evidence that ESG ratings are related to stock returns in our global sample from 2001-2020. Of the ten ESG ratings considered in Table 1, only one rating (ISS) has a coefficient that is statistically significant (point estimate of 0.007, significant at the 10% level). For the other ratings, coefficients vary between -0.006 and 0.003 (thus taking on both negative and positive values) and are not significant at conventional significance levels. The effect size (economic significance) of the ESG rating coefficients is also generally small. The largest coefficient in absolute value (0.007 for ISS) indicates that a one sample standard deviation increase in *ESG* is associated with a relatively modest additional stock return of 1.26% per annum.[5] Effect sizes for the other ratings tend to be considerably smaller. Not only are the point estimates of the effect sizes small, but the upper bounds of the 95% confidence intervals of the

[5]Even this effect size based on the largest (and the only significant) ESG coefficient in Table 1 is notably smaller than, for example, the additional stock return of 1.8% to 4.0% per annum (depending on the specification) associated with a one standard deviation increase in carbon emissions documented by Bolton and Kacperczyk (2021).

coefficient point estimates further tend to suggest that large effect sizes are unlikely. For example, considering a one standard deviation shock to each ESG rating, we can formally reject the hypothesis that stock returns increase by 1.8 percentage points or more per annum for seven of the ten raters and by just one percentage point or more for three of the ten raters. For the three raters for which we cannot reject the hypothesis that stock returns increase by 1.8 percentage points or more (ISS, FTSE, and *Composite 6+*), the upper bounds of the confidence intervals vary between 2.4% and 2.9%. The fact that the effect sizes for these three raters are less precisely estimated (and thus the confidence intervals are larger) may be a consequence of their shorter time-series relative to other raters. The results are similar if we look at the lower bounds of the confidence intervals. These findings suggest that the very weak statistical evidence in Table 1 is not *just* due to limited statistical power to detect an economically meaningful effect size, but also due to the small ESG coefficient point estimates.[6] Section 3.2 below provides a more detailed analysis of effect sizes and the statistical power of our tests.

The lack of evidence on a relation between ESG and stock returns in our global sample is striking in light of the large number of studies suggesting a positive relation (Friede et al. 2015). We also note that averaging ESG ratings across different raters (as in our *Composite 3+* and *Composite 6* ratings) does not seem to be an effective way to reduce the noise in a potential "ESG signal" about future stock returns. One alternative interpretation of Table 1 is that the ISS ESG rating *does* have a reliably positive association with stock returns around the world. However, we note that our statistical tests do not account for the multiple hypothesis testing problem (Harvey et al. 2016, Heath et al. 2023b), and that finding one significant ESG rating coefficient at the 10% level in ten tests, under the null hypothesis that the ESG rating coefficient is equal to zero, is exactly what one would expect if the test is well-specified.

Of course, it is possible that the results would be stronger under different empirical specifications, or for specific subsamples of our large global sample. To assess this possibility, Figure 3 (Panels A–J) graphically presents the results of ten different variations of the baseline regressions presented

[6]The coefficients on the control variables in Table 1 are largely insignificant – except size, profitability, and R&D in a number of the regression models – in line with Green, Hand, and Zhang (2020), who find that only two out of the 94 stock characteristics they consider independently predict stock returns. Gibson et al. (2021), who also study ESG ratings, similarly find little evidence of stock return predictability based on stock characteristics. The strongest predictor in Table 1 is R&D, consistent with Hou, Hsu, Wang, Watanabe, and Xu (2022). In unreported analyses, we find similar results on the ESG-return relation when we drop R&D as control variable.

in Table 1. Each panel of Figure 3 shows the point estimates (in bars) – with associated 95% and 99% Newey-West confidence intervals – of the coefficients on the ten different individual and composite ESG ratings from Table 1. In other words, each bar represents the ESG rating coefficient from a separate regression that deviates in one dimension from the baseline regression specification in Table 1 (but with the same control variables).

An issue that may be relevant is that ESG ratings exhibit strong country- and industry-level components (Gillan et al. 2021), which raises the question of whether a firm's ESG performance should be evaluated globally or relative to its country and/or industry peers. Panel A of Figure 3 shows the results when we country- and/or industry-adjust by demeaning ESG ratings each month at the country-, industry-, or country- and industry-level. We use six-digit Global Industry Classification Standard Codes (GICS) and assign stocks to countries based on the location of the stock exchange where the stock is listed. For comparison, Panel A also shows the baseline results with unadjusted ratings.[7] The bottom line is that these adjustments do not materially alter the magnitude or statistical significance of the coefficients. Out of the 30 coefficients on the adjusted ESG ratings in Panel A, only the one on the country-adjusted ISS rating is significant at the 5% level or better.

Panel B of Figure 3 shows the ESG rating coefficient estimates of our baseline regressions estimated separately for the E, S, and G subratings. Such subratings are not available for the *RRR* rating from RepRisk. None of the 27 subrating coefficients in Panel B is statistically significant, and the effect sizes, if anything, are smaller in absolute value than those in Table 1. In unreported analyses, we reach the same conclusion when we country- and/or industry-adjust the subratings.

We further estimate our regressions separately for different major geographic regions: Asia-Pacific, Emerging Countries, Europe, Japan, and North America. The results are in Panel C of Figure 3. The finding that there is little evidence that ESG ratings are informative about future stock returns also holds for every major region. Only seven out of the 50 ESG coefficients in Panel C are significantly positive at the 5% level: *RRI* for Asia-Pacific, *Composite 3+* for North America, *Composite 6* for Emerging Countries and Europe, *ISS* for Europe and North America, and *S&P* for Europe. In addition, *S&P* is significantly negative for Japan. Given the large number of hypotheses

[7]For expositional reasons, Figure 3 does not show 90% confidence intervals, which is why the coefficient on the non-adjusted ISS ESG rating (significant at the 10% level in Table 1) appears insignificant in Panel A of Figure 3.

tested in this panel of Figure 3 in particular (and the relatively high number of individual rejections – eight – of the null hypothesis of no ESG-return relation), we follow the recommendation of Heath et al. (2023b) and apply the correction of Benjamini, Krieger, and Yekutieli (2006) which produces sharpened false discovery rate (FDR) q-values.[8] Based on this adjustment, all 50 coefficients turn insignificant at conventional levels. In light of the multiple hypothesis testing problem, and given the lack of consistency regarding which rating matters for which region, in our view the only appropriate conclusion from Panel C is thus that there is no reliable evidence of a relation between ESG ratings and stock returns across these five major regions. In unreported analyses, we obtain similar results when we country- and/or industry-adjust the ratings, or use the E, S, and G subratings instead.

Next, we examine the conjecture that the relation between ESG ratings and stock returns may be stronger in recent years in which attention to ESG considerations has increased. Panel D of Figure 3 shows the coefficient estimates from our global regressions (using non-adjusted, country- and/or industry-adjusted ratings) estimated for the subperiod 2016-2020 that starts after the signing of the Paris Climate Agreement in December 2015. We choose this subperiod for two reasons. First, the Paris Agreement likely raised investor and consumer awareness of sustainability issues (Bolton and Kacperczyk 2021, Klausmann, Krueger, and Matos 2024). Second, in this subperiod, all ESG rating agencies we consider cover a substantial number of stocks, thus potentially marking a greater maturity in the ESG rating industry, and allowing for a more direct comparison of results across raters. The results are similar to those presented in Panel A for the full sample period 2001-2020. Out of 40 ESG rating coefficients, only two are significantly positive (country-adjusted ISS and country- and industry-adjusted $MSCI$) and one is significantly negative (country- and industry-adjusted RRR) at the 5% level. In unreported analyses, we rerun this exercise by region and find similar results. For example, for the unadjusted ratings, we find that eight of the 50 coefficients are significant at the 5% level, one of which is also negative. Once we adjust this result for multiple hypothesis testing using the correction of Benjamini et al. (2006), all 50 coefficients become insignificant.

The paucity of evidence of a relation between the E rating and global stock returns may be surprising in light of Pástor et al. (2022), who find that green stocks (identified using the MSCI

[8] We note that, unlike other adjustments for multiple hypothesis testing, this method can produce sharpened q-values that are smaller than the initial p-values and has comparatively greater statistical power than other alternatives.

E rating) outperformed in recent years (the "greenium"). To reconcile these seemingly conflicting findings, we rerun our analysis starting in November 2012 as in Pástor et al. (2022). Since they study the U.S., use different control variables, and use a transformed version of the MSCI E rating, we run separate analyses for the U.S. and the rest of world (RoW), use two different sets of control variables, and use both our MSCI E rating (MSCI I) and the E rating of Pástor et al. (2022) (MSCI II).[9] The results are in Panel E of Figure 3. We replicate their key result: the coefficient on both versions of the MSCI E rating is positive and significant for the U.S. when we control for the stock characteristics corresponding to the Fama and French (2015) five-factor model plus momentum. Here, we also find a significantly positive coefficient for the ISS and *Composite 3+* E ratings. With more elaborate controls, only the ISS E rating remains significant, whereas the *Composite 3+* E rating is now marginally significant and the MSCI E rating becomes either insignificant or marginally insignificant depending on the version of the rating used. For the RoW, none of the ten E ratings is significant, regardless of the controls used – in line with a contemporaneous paper by Karolyi et al. (2023), which concludes that the global greenium effect they document using the MSCI E rating "mostly occurs in North America and during the period before 2016."

Panel F of Figure 3 shows the results of our baseline analyses when we substitute the *level* of the ESG ratings as in Table 1 with ESG *momentum* (Bekaert et al. 2023), which we define as the most recent change in ESG ratings during the previous 12 months (using non-adjusted, country- and/or industry-adjusted ratings). Out of the 40 coefficients on ESG momentum in Panel F, only one is significantly positive (non-adjusted *Composite 3+* momentum) and one is significantly negative (country- and industry-adjusted S&P momentum). In short, we also find very little evidence that ESG momentum helps explain the cross-section of stock returns. In unreported analyses, we obtain similar results when we focus on different regions or E, S, and G subratings.

Prior studies suggest that the stock market may respond differently to ESG rating upgrades and downgrades (Krueger 2015, Shanaev and Ghimire 2022). To examine such potential asymmetries in the relation between ESG ratings and global stock returns in our sample, Panels G and H of Figure 3 show the results for downgrades and upgrades, respectively, based on our various country- and/or industry-level adjustments. We capture downgrades (upgrades) by individual raters with dummy

[9] Pástor et al. (2021) compute the E rating as $-\frac{(10-E)\times E_{weight}}{100}$ where E_{weight} is the firm-specific weight assigned to the E rating in the computation of the overall ESG rating.

variables that equal one if ESG momentum is negative (positive) for a given stock-month. For the composite raters, we measure downgrades (upgrades) as the proportion of raters that downgraded (upgraded) a given stock out of the total number of raters that rated the stock. We find that ESG downgrades are unrelated to stock returns, with only two coefficients out of 40 coefficients significant at the 5% level with negative sign: unadjusted MSCI and unadjusted *Composite 3+*. The results for upgrades in Panel H are even more unequivocal: all coefficients are statistically insignificant. In unreported analyses, we find similarly weak results when we run separate analyses by geographic region, when we use E, S, and G subratings instead of ESG ratings, and when we restrict the analysis to the second half of the sample period when there are more upgrades and downgrades.

Next, we examine best-in-class and exclusion strategies. These ESG strategies are widespread among investors (Amel-Zadeh and Serafeim 2018) and consist of selecting stocks with the best (or dropping stock with the worst) ESG rating relative to their peers. We indicate best-in-class (worst-in-class) stocks with a dummy variable that equals one for stocks in either the top (bottom) 10% or 20% of the distribution of ESG ratings in their sector in a given month. If best-in-class (exclusion) strategies boost financial returns, we would expect positive (negative) coefficients on the best-in-class (worst-in-class) dummies. Panels I and J of Figure 3 show the results for the global sample and for the U.S., respectively. We show separate results for the U.S. because, to our knowledge, most studies that find evidence that best-in-class in strategies outperform use U.S. data (Kempf and Osthoff 2007, Statman and Glushkov 2009). The results show that all coefficients in Panel I (global tests) are statistically insignificant and that all but one coefficient in Panel J (U.S. tests) are also insignificant. In unreported analyses, we obtain similar results when we rerun these analyses based on ESG (sub)ratings in each region separately.[10]

Our choice for Fama-MacBeth regressions in our baseline analyses was motivated by the fact that these regressions (i) take an investor perspective since the ESG rating coefficients each month can be interpreted as the monthly returns on a zero-investment portfolio that is invested according to the ESG rating signal and (ii) enable straightforward controlling for other well-known cross-sectional return predictors that could potentially explain any ESG rating effect. That said, Fama-MacBeth regressions have as a drawback that the results may be disproportionally driven by microcaps,

[10] We do not report results for RepRisk in Panels I and J because there is not enough variation in the RepRisk data to reliably define best- and worst-in-class within each sector-month. This happens because many firms do not have regular risk incidents captured by the media, thus receiving the same score.

which represent a tiny fraction of the market in terms of market capitalization. Therefore, we re-examine our baseline analyses based on value-weighted Fama-Macbeth regressions as well as based on portfolio sorts.

For the value-weighted Fama-Macbeth regressions, we use lagged market capitalization to form weights and re-estimate all regressions in Panels A to J of Figure 3. The results, which are reported in the Internet Appendix, are qualitatively similar to our baseline findings. To facilitate comparison, we present equally-weighted and value-weighted results side by side.

For the portfolio sorts, we sort stocks at the end of each month into high and low ESG rating portfolios and compute the monthly return spread (high minus low) in excess of the one-month U.S. Treasury bill rate. In computing portfolio returns, we value-weight stocks based on their market capitalization at the end of the previous month. We then estimate alphas of these portfolio strategies relative to different factor models by running time-series regressions of the value-weighted return spreads on the factor returns for each factor model. We conduct the portfolio sorts separately for all combinations of geographic regions, ESG (sub)ratings, country and/or industry rating adjustments, and three alternative portfolio breakpoints (top and bottom 10%, 20%, or 30%). We examine the raw return spread (no controls) as well as the alpha relative to five different factor models: the CAPM, the Fama and French (1993) three-factor model (*FF3*), the Carhart (1997) four-factor model (*Carhart*), the Fama and French (2015) five-factor model (*FF5*), and the Fama-French five-factor model plus momentum (*FF5+MOM*)).

The results are reported in Table 2, where each column corresponds to a specific combination of ESG (sub)rating and country and/or industry adjustment, and each row represents a combination of a factor model and geographic region. For example, column (1) shows the results using the overall ESG ratings without country and/or industry adjustments, broken down by factor model and geographic region. As reported in the bottom row, this column reports the results for 810 different portfolio sorts (5 regions, 6 factor models, 3 breakpoints, and 9 ESG ratings).[11] For brevity, the table does not report the individual alpha estimates for the total of 11,920 portfolio sorts. Instead, it reports two numbers "x/y", where x is the number of sorts with a statistically significant alpha at the 5% level or better, and y indicates how many of those significant alphas are positive. For

[11] We exclude RepRisk's RRI from the analysis due to limited cross-sectional variation across firms for sorting purposes.

example, a hypothetical entry “3/2” indicates that, out of all sorts for that specific combination of (sub)rating, rating adjustment, region, and factor model, three alphas are statistically significant, and two of those are positive. The total number of sorts per combination and per column is indicated at the bottom of the table. Given the very large number of tests in this table, significance tests are corrected for multiple hypothesis testing using the procedure of Benjamini et al. (2006), applied separately per column (i.e., for each (sub)rating and each rating adjustment).

The key takeaway from Table 2 is that portfolio sorts also yield very little evidence supporting a relation between ESG ratings and stock returns. The vast majority of entries are 0/0. The few exceptions do not seem to point to robust occurrences of a significant relation between ESG ratings and stock returns in a certain category of sorts, since the significant results still are a minority of the total number of sorts in that category and since the significant alphas almost always disappear in the factor models that include more factors. For example, in column (9), we observe that a statistically significant and positive alpha is found using country-adjusted ESG ratings for Emerging Countries in either 2 or 4 sorts (out of a total of 27 sorts in that category), depending on whether we use the CAPM or the Fama-French three-factor model. In other words, there are some combinations of portfolio breakpoints, factor models, and ESG ratings that show significant and positive alphas for Emerging Countries. However, this finding is not robust across alternative factor models and only holds for specific combinations of portfolio breakpoints and ESG raters. In the Internet Appendix, we demonstrate that we obtain similar results for equally-weighted portfolios and when the analysis is limited to start in November 2012, as in Pástor et al. (2022).

Overall, the evidence in Tables 1 and 2 and in Figure 3 indicates that the relation between ESG ratings and global stock returns over 2001-2020 is insignificant. This result obtains across different rating agencies, regions, time periods, ESG (sub)ratings, ESG momentum, ESG downgrades and upgrades, best-in-class and exclusion strategies, and using different ways to adjust for country and/or industry effects in ESG ratings. In unreported analyses, we show that our results are also qualitatively unaffected if we drop financial stocks, run regressions by sector, or use a less elaborate set of data filters. Although we cannot rule out that there may be more subtle patterns in the relation between ESG ratings and stock returns that our analyses do not detect, we believe that our battery of tests revealing no relation does suggest that any potentially non-negligible ESG-return relation is limited to very specific settings and requires a compelling justification and careful consideration of the multiple hypothesis testing problem.

In the next four subsections, we investigate whether accounting for, respectively, statistical power, ESG uncertainty, country characteristics, and measurement error can shed more light on the ESG-return relation.

3.2 Effect sizes and statistical power

The analyses in the previous subsection show that we rarely reject the null hypothesis of no relation between ESG ratings and future stock returns. As discussed above, this failure to reject the null hypothesis could stem from either a weak association between ESG ratings and future stock returns in our sample or from insufficient statistical power to identify such effects. In this subsection, we provide a more elaborate analysis of effect sizes and statistical power.

In the previous subsection, our discussion of effect sizes was limited to the baseline regressions in Table 1. In Table 3, we report summary statistics for the effect sizes (based on the ESG rating coefficient point estimates) for all regressions shown in Panels A through J of Figure 3, disaggregated by panel. The table shows the mean, median, and 25^{th} and 75^{th} percentiles of the effect size distribution as well as the median lower and upper bounds of the 95% confidence intervals of the effect size point estimates for each panel. For ease of interpretation, we express effect sizes and confidence interval bounds as annualized stock return responses (in %) to one standard deviation shocks to ESG/E/S/G (sub)ratings in Panels A through F. In Panels G through J, which show results for ESG upgrades, downgrades, and best/worst-in-class, we also express estimates as annualized stock returns responses (in %) but we do not use standard deviation shocks because the predictors are dummy variables.

The results in Table 3 show that the effect size distributions are centered around zero for most of the panels in Figure 3. Furthermore, the mean, median, and 25^{th} and 75^{th} percentiles of the effect size distribution tend to be relatively small, often (considerably) below 1% in absolute value. For some panels, the effect size distribution is not centered around zero and the percentiles of the effect size distribution are somewhat larger in absolute value, but still these patterns do not seem to point at consistently large effect sizes for these subsets of our empirical tests. For example, the largest positive median effect size occurs for Panel E, which shows the relation between stock returns and E subratings in the U.S. and the rest of the world. The median (mean) effect size for this panel has a non-negligible but still modest economic magnitude of 0.958% (1.048%) per annum and the 75$^{\text{th}}$ percentile is below 1.25%.

The confidence interval bounds reported in Table 3 by-and-large paint a similar picture. Based on the median lower and upper bounds, the confidence intervals tend to be centered around zero for each panel. For many panels, albeit not all, the confidence intervals are narrow enough to formally rule out (that is, statistically reject) medium to large effect sizes. For example, the results for the lower and upper bounds in columns (5) and (6) indicate that we can reject the null hypothesis that the effect size is 1.9% per annum or larger (in absolute value) in eight of the 12 panels, for half the estimates in those panels (since these are the median lower and upper bounds). That said, for some panels, our tests do not allow us to formally rule out potentially considerable effect sizes of 2.5% to 3.5%. This holds in particular for Panel E (E subrating) and for Panels I and J (best/worst-in-class).

Taken as a whole, the findings in Table 3 suggest that the relation between ESG ratings and stock returns is likely to be genuinely weak across a large number of different combinations of ESG (sub)ratings, regions, and sample periods. However, the individual effect sizes are not always estimated with sufficient precision to definitively rule out, through formal hypothesis testing, the possibility that effect sizes could sometimes be larger. This raises a natural follow-up question: what is the magnitude of effect sizes that we have sufficient statistical power to reliably detect? In other words, how concerning is the magnitude of potential type II errors in our analysis?

To address this question, we examine the minimum detectable effect size (MDES) of Bloom (1995) in each of our tests. The MDES quantifies the smallest effect size that, if true, we could reliably detect given our data and research design. In line with Bloom (1995), we compute the MDES for each of our individual tests as the smallest true effect size we can detect with 80% statistical power at the 5% significance level, using a two-sided hypothesis test. Several recent finance papers use the MDES to gauge the power of statistical tests (Coles, Heath, and Ringgenberg 2022, Simeth and Wehrheim 2024). Table 4 reports the the mean, median, and 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles of the MDES distribution, separately for each panel of Figure 3. The MDES can be expressed in the same units as the regression coefficient to which it refers, but it is common to consider two different ways of expressing the MDES: relative to the standard deviation of the dependent variable as well as in absolute terms. Following this approach, Panel A of Table 4 presents the MDES expressed as a fraction of one standard deviation of monthly stock returns in our sample, and Panel B presents the

MDES expressed as annualized returns (in %). In both panels, we compute effect sizes relative to a one standard deviation shock to an ESG rating (since the economic magnitude of our regression coefficients are harder to interpret directly), unless the ESG rating is defined as a dummy variable (i.e., upgrades/downgrades and best/worst-in-class).

Panel A of Table 4 indicates that, on average, we can detect effect sizes that are 41 to 208 times smaller than the sample standard deviation of monthly stock returns. Additionally, in each panel, all MDES that fall within the interquartile range are also small (at least 27 times as small as the sample standard deviation). This suggests that, relative to the standard deviation of monthly stock returns, our tests have substantial statistical power to detect a relation between ESG ratings and stock returns. One way of viewing this is that, given that stock returns are our dependent variable, the statistical power of our tests to detect effect sizes that are large in light of the distribution of stock returns (that is very dispersed around the mean) is substantial. That said, the fact that stock returns are inherently noisy may still imply that we are not able to detect effect sizes that are economically large from an investor's perspective.

For a more complete picture of the MDES, we thus also examine the MDES expressed in annualized returns (in %) in Panel B of Table 4. Here, the conclusion is more nuanced. The median annualized MDES in column (3) ranges from 0.65% to 2.3% in eight of the ten panels, with a maximum of 3.4% for Panel I (global best/worst-in-class). These numbers indicate that we have sufficient power to detect medium to large effects in a wide range of tests (albeit not all), and that our tests on various occasions may lack power to detect more modest, albeit not negligible, effect sizes. The 25^{th} percentiles of the MDES distribution in column (2) confirm this; these percentiles are between 1 and 2% for most of the panels, which indicates that our tests are often not powerful enough to detect effect sizes smaller than these numbers. And the 75^{th} percentiles of the MDES distribution in column (4) indicate that, on some occasions, our tests are unable to detect considerable effect sizes of 2 to 3% or even higher. The percentiles of the MDES distribution tend to be the largest for Panel C (different regions), Panel E (E subrating) and Panels I and J (best/worst-in-class). It is perhaps not surprising that the power of our tests is relatively lower on these occasions, as they are based on geographic subsamples (Panel C), potentially noisier subratings and the U.S. subsample only (Panel E), and dummy variables that throw away information contained in the continuous ESG rating (Panels I and J). We note that for the baseline tests in Panel A of Figure 3 that are based on our full sample, statistical power seems to be relatively strong, with a median MDES of 1.405%.

The findings in this subsection do not show that statistical power is never a concern in any of our tests. But, taken together, we conclude that a lack of statistical power is unlikely to explain the general absence of a relation between ESG ratings and future stock returns over the past two decades. Although we do not rule out that there may be combinations of ratings, regions, and sample periods for which a stronger relation could exist that our tests may fail to detect, our tests generally have enough statistical power to identify medium to large effects across a broad range of tests. Moreover, the fact that we systematically fail to identify such larger effects in the actual data across a wide variety of specifications (i.e., all panels of Figure 3) indicates that the general failure to reject the null of no relation between ESG ratings and stock returns in our paper is likely due to the relation truly being weak.

3.3 Is the ESG-return relation stronger when there is less ESG uncertainty?

Avramov et al. (2022) argue that uncertainty about a firm's "true" ESG performance may have a bearing on the relation between ESG and stock returns. In particular, their model and empirical evidence for the U.S. stock market suggest a negative relation between ESG and stock returns – but only when ESG uncertainty is low. Such a negative ESG-return relation is consistent with the notion that investors' non-pecuniary preferences for ESG may result in lower expected returns for stocks with stronger ESG performance, in line with the theoretical models by Fitzgibbons et al. (2021) and Pástor et al. (2021). Intuitively, ESG rating disagreement across different raters leads investors to be uncertain about stocks' ESG performance, thus impeding their ESG preferences from being incorporated into stock prices.

In this subsection, we examine the possibility that our finding of an insignificant ESG-return relation so far changes when taking into account ESG uncertainty – which could arguably be greater in our global sample than in the U.S. To this end, we follow Avramov et al. (2022) in measuring stock-level ESG uncertainty as well as average ESG. First, each month, we convert each of our ESG ratings to percentile ranks. Second, each month, we compute stock-level ESG uncertainty as the standard deviation of these percentile ranks across the different ratings. Third, each month, we compute stock-level average ESG as the average percentile rank across the different ratings, using a minimum of two different ratings (*Composite 2+*). We exclude RepRisk ratings from ESG uncertainty as well as from *Composite 2+* because RepRisk ratings are negatively correlated with other raters, which would mechanically generate higher disagreement for stocks covered by RepRisk. In line with Avramov et al. (2022), we further define a dummy variable (*Low ESG uncertainty*) that equals one if a stock is among the 20% of stocks with lowest ESG uncertainty in a given month and region, and zero otherwise.

In the spirit of Avramov et al. (2022), we then run Fama-Macbeth regressions of global stock returns on *Composite 2+* and *Low ESG uncertainty*, as well as the interaction of these two variables. The hypothesis of interest is that the interaction term has a negative coefficient. We present the results in Table 5. We show separate results for the pre-2013 (Panel A), post-2013 (Panel B), and full sample period (Panel C), because Avramov et al. (2022) find that the interaction effect is not significant in the most recent decade, which they attribute to the possibility that unexpected demand effects obscure the equilibrium relations predicted by their model in recent years. Each Panel of Table 5 shows the coefficient estimates of *Composite 2+*, *Low ESG uncertainty*, and the interaction term for each of the five major geographic regions, for the global sample, and – to facilitate comparison with Avramov et al. (2022) – also separately for the U.S. We include the same control variables as in Table 1, but their coefficients are suppressed to conserve space.

We confirm the key finding of Avramov et al. (2022) that the coefficient on the interaction term is significantly negative for the U.S. (and also for North America) in the pre-2013 period but not in the post-2013 period. Thus, we find evidence of a negative ESG-return relation for stocks with low ESG uncertainty for the U.S. in the first half of the sample period. But there is no evidence of a significantly negative interaction effect for any of the other regions or for the global sample, neither pre-2013 nor post-2013. For Asia-Pacific pre-2013 and for Japan post-2013, we instead find a significantly positive interaction effect.

Overall, the evidence in Table 5 suggests that accounting for ESG uncertainty adds little to our understanding of the insignificant relation between ESG and global stock returns.

3.4 Does the ESG-return relation depend on country characteristics?

So far, our analyses have aggregated countries to the regional or global level, but the literature has put forward various hypotheses on why and how the relation between ESG and stock returns could exhibit heterogeneity across individual countries. In this subsection, we assess three different cross-country hypotheses on the ESG-return relation (concerning country-level ESG social norms, ESG disclosure standards, and ESG regulations). A further motivation for examining these hypotheses is the argument of Heath et al. (2023b) that testing additional hypotheses may improve inference about the main relation of interest, especially in the face of the multiple hypothesis testing problem. For each hypothesis, we briefly discuss the motivation from the literature as well as the empirical proxies we use to test it, before turning to the results of these tests. We refer to the Internet Appendix for a detailed description of the variables used.

3.4.1 ESG social norms

A key reason why ESG ratings could be related to stock returns is that investor preferences for stocks with higher ESG ratings could result in (i) a positive ESG-return relation in the short term as demand effects lead to these preferences being priced in (Pástor et al. 2022) or (ii) a negative ESG-return relation in equilibrium (Fitzgibbons et al. 2021, Pástor et al. 2021). The ESG-return relation could thus be stronger in countries with social norms reflecting more positive attitudes and beliefs regarding ESG issues, as investors in these countries may have stronger ESG preferences (Dyck et al. 2019). These social norms can also play a role via a customer channel if stronger ESG firms can increase sales more in countries where customers value ESG more (Aghion, Bénabou, Martin, and Roulet 2023) and if these effects are not fully anticipated by financial markets.

We measure country-level ESG social norms in three different ways. First, following Krueger et al. (2024), we construct two indices of social and environmental norms which aggregate the responses to various survey questions in the Integrated Values Survey. Second, we use the indices of social movement activity and associational activity developed by Welzel (2013), which measure the extent to which individuals in a country are involved in social movements and recreational, humanitarian, and environmental organizations, respectively. Third, we measure the political orientation of a countries' citizens based on whether their voting preferences and the political parties in power lean towards the left or the right of the political spectrum, using various variables constructed based on the Comparative Political Data Set.

3.4.2 ESG disclosure standards

Countries vary widely in the strictness of ESG disclosure standards (Krueger et al. 2024), possibly leading to a poorer ESG information environment in some countries. Survey evidence indicates that institutional investors consider data quality a key challenge in ESG investing (BNP Paribas 2023, Ilhan et al. 2023). To the extent that stricter disclosure standards lead to the production of higher-quality ESG information, ESG ratings may be more strongly related to stock returns in countries with stricter ESG disclosure standards, because ESG ratings incorporate more value-relevant information (Fitzgibbons et al. 2021) and/or because of a reduction in ESG rating uncertainty (Avramov et al. 2022).

To measure country-level ESG disclosure standards, we follow Krueger et al. (2024) and exploit the fact that several countries in our sample have passed mandatory ESG disclosure regulations at different points during our sample period, while others did not (thereby creating a natural control group). We define a dummy variable that equals one if a country has mandatory ESG disclosure regulations in place at a given point in a time, and zero otherwise. We define similar variables for the E, S, and G dimensions of ESG. Using the data compiled by Krueger et al. (2024), we further use variables that isolate mandatory ESG disclosure (i) on a full-compliance basis as opposed to comply-or-explain basis, (ii) mandated by government as opposed to other entities such as stock exchanges, and (iii) mandated all at once for the three dimensions of ESG.

3.4.3 ESG regulations

The strictness of a country's regulations on ESG issues (beyond disclosure) could also affect the strength of the ESG-return relation. For example, in countries with stricter environmental regulations, investors may be more concerned about climate transition risks (Bolton and Kacperczyk 2023). Along the social dimension, Edmans et al. (2024b) find that the link between employee satisfaction and stock returns is stronger in countries with less strict labor market laws, in which the value of employee satisfaction (not immediately incorporated into stock prices) is greater.

We measure the strictness of environmental regulations with the OECD environmental policy stringency index, the Yale University environmental performance index (which measures the performance of government environmental policy), and the Access Initiative / World Resources Institute environmental democracy index. These measures are widely used in the literature (Dyck et al. 2019, Martínez-Zarzoso and Morales-Lage 2019). To measure the strictness of social regulations, we use the employment laws index of Botero, Djankov, Porta, de Silanes, and Shleifer (2004), the labor regulation index of the Fraser Institute, and two versions of the OECD employment protection legislation used by Edmans et al. (2024b).

3.4.4 Tests of the three cross-country hypotheses

In the spirit of Edmans et al. (2024b) and Bolton and Kacperczyk (2023), we test these three cross-country hypotheses by running panel regressions of stock returns on lagged ESG ratings, country characteristics, and the interaction of the ESG ratings with the country characteristics. We include one country characteristic and one interaction term per regression. We run each regression separately for each of the ten ESG ratings from Table 1. We control for the full set of stock characteristics used in Table 1 as well as country-month fixed effects. Standard errors are double clustered at the month and stock levels. We start our sample in January 2014 to ensure sufficiently broad ESG rating coverage. The hypotheses each predict significant coefficients on the interaction terms.

We summarize the results in Table 6. For brevity, the table does not report coefficient estimates, but rather shows, for each country characteristic associated with one of the three hypotheses, the number of coefficients on the interaction term (out of a maximum of ten coefficients corresponding to the ten individual and composite ESG ratings considered) that are statistically significant at the 5% level, as well as the number of those significant coefficients that are positive. Panels A, B, and C show these numbers for the hypotheses on, respectively, ESG social norms, ESG disclosure standards, and ESG regulations. In each panel, columns (1) through (4) show these numbers for, respectively, the ESG, E, S, and G ratings. Column (5) shows the number of countries for which each country characteristic is available. To illustrate the information presented, a hypothetical entry "5/3" in Table 6 would mean that, out of the ten different ESG ratings, five exhibit a statistically significant interaction effect with a given country characteristic, and that three of those five significant interaction coefficients are positive.

The main takeaway from Table 6 is that there is very little evidence that any of the 29 country characteristics we used to test the three cross-country hypotheses significantly moderates the ESG-return relation. Across all country characteristics and across the ESG, E, S, and G ratings, at most one – and more often zero – out of the maximum of ten interaction terms (based on the ten different ESG ratings) has a coefficient that is significant at the 5% level. We thus find little support for the three hypotheses on how country characteristics could affect the ESG return-relation. These results suggest that the lack of a relation between ESG and stock returns holds globally, irrespective of cross-country differences in ESG social norms, ESG disclosure standards, and ESG regulations.

3.5 Measurement error

As discussed above, a potential concern is that ESG ratings may measure the "true" ESG only with measurement error – an issue we discuss in more detail in this subsection. If this was the case, it would lead to attenuation bias and reduced statistical power. Although we do not contest that the ESG ratings we use could contain noise, our analyses try to mitigate this concern in four ways. First, we show that our results are consistent when using two different composite ratings next to the individual ratings. In Figure 3, ratings from individual data providers often exhibit substantially larger (though statistically insignificant) coefficients compared to composite ratings, which suggests that attenuation bias is unlikely to play a major role. Second, the evidence in the previous subsection that a substantial number of our tests have sufficient statistical power to detect economically meaningful effects suggests that measurement error is unlikely to undermine our conclusions by weakening the power of our tests. Third, we use seven major raters and adjustments that are widely used by investors and across the board fail to find a robust association between ESG ratings and stock returns. This helps us rule out that our results are driven by some ratings being particularly noisy or by raters that revise their ratings. Fourth, our findings on ESG uncertainty in Section 3.3 indicate that even composite ratings are rarely predictive of future stock returns, even in cases where disagreement among raters on ESG ratings is the lowest. If cross-rater disagreement reflects noise in the rating process, one would expect ratings with low cross-rater disagreement to be less noisy and thus more predictive of stock returns. Our results do not support this interpretation.

Moreover, there is a conceptual reason to be skeptical that measurement error is a major concern in our setting. Measurement error assumes the existence of a unique "true" value of ESG that ratings aim to capture, with significant disagreement among raters potentially reflecting raters' inability to quantify this "true" value of ESG. However, the subjective and multidimensional nature of ESG complicates this notion, as there is no universally accepted definition of ESG. In our view, it is hard to conceptualize the existence of a "true" value of ESG given the subjectivity and multidimensionality of the concept. As pointed out by Starks (2023), concepts of ESG vary across investors, ranging from investors who view ESG through the lenses of non-pecuniary preferences and the impact of firms' on stakeholders (*values* approach), to investors that take the view that ESG considerations affect firms' financial value (*value* approach). This suggests that there is not a uniquely "true" value of ESG. What could be seen as measurement error from the perspective of one investor might be considered "true" ESG from another investor's perspective. In line with this view, Eccles and Stroehle (2018) document that raters rely on different concepts of ESG depending on whether they have a value or values approach.

Given the absence of a universally true value of ESG, it may be reasonable to think of raters as giving opinions about the ESG profile of firms in the form of ratings based on different views about which ESG attributes are relevant, which data/variables best serve the purpose of measuring those attributes, and how to weigh those attributes (Edmans 2023). If one takes this view, statistical techniques to remove measurement error may have the side effect of removing, or under-weighting the variation in ESG ratings that corresponds to the unique opinions of each rater. It is, in our view, theoretically unclear whether or not this is a promising avenue when studying the relation between ESG ratings and future stock returns. For example, the most obvious metrics of ESG that investors can agree to be "true" and relevant are exactly those that are the most likely to be priced in by the market. In the limit, if there was only one obviously correct ESG rating, it would likely be priced in by the market and would not be useful to predict future stock returns. A similar argument can be made for a composite rating that captures the common variation across major raters whose ratings are easily available to investors. Moreover, since different raters rely on different concepts of ESG, it is unclear whether the application of statistical techniques to remove measurement error results in a better measurement of ESG or in a new ESG measure which combines disparate ESG concepts without strong theoretical motivation for doing so.

Nevertheless, we acknowledge that the number of possible combinations of ESG metrics is extremely large, especially if one also considers non-major ESG raters, ESG rating data disaggregated at a granular level, ESG data other than ratings, and/or sophisticated methods to combine ESG metrics. Given that, it is conceivable that there may be some combinations that are informative about future cash flows and that the market has not priced in those combinations yet. It is also possible that combining ESG ratings does not reduce measurement error stricto sensu but helps capture demand for ESG stocks by better capturing pro-social investors' overall ESG preferences or ESG risk concerns, thus being more likely to capture cross-sectional variation in expected returns. If this is the case, we would expect to observe a negative relation between expected returns proxies, such as the implied cost of capital, and carefully constructed combinations of ESG metrics that capture societal preferences for ESG. However, any such evidence should be carefully interpreted in light of the multiple hypothesis testing problem and the number of possible combinations of ESG metrics being extremely large. Exploring these issues in depth is an interesting avenue for future research.

3.6 Discussion of our findings relative to the literature

Whether and how stock returns depend on ESG (ratings) is still widely debated. Practitioners tend to hold the view that ESG is generally associated with higher risk-adjusted performance. Evidence of this can be found in FT (2017) or the recent survey by Edmans, Gosling, and Jenter (2024a), which finds that almost half of the fund managers surveyed associate better environmental and social performance with positive alpha. In contrast, academic thinking on the question is often more skeptical and stresses the opposite relation, i.e., a negative association between ESG ratings and expected stock returns in equilibrium.

When taking a bird's eye view of the vast literature on the stock return implications of ESG, there is evidence in favor of a positive, negative, or no relation at all. For instance, early studies such as Hamilton, Jo, and Statman (1993) find that there is no relation between ESG and stock returns. In contrast, Kempf and Osthoff (2007), Statman and Glushkov (2009), and Edmans (2011), or most recently Pástor et al. (2022) present evidence of a positive relation. Some researchers also find a negative relation between ESG and stock returns (Renneboog, Ter Horst, and Zhang 2008, Geczy, Stambaugh, and Levin 2021). Given the variety and diversity of results in the literature, meta studies, such as Friede et al. (2015) or more recently Atz, Van Holt, Liu, and Bruno (2023), have attempted to aggregate the vast body of studies. These meta studies typically find that most studies are more likely to document non-negative (i.e., insignificant or positive) relations between stock returns and measures of ESG.

A possible reason for the large spectrum and divergent results in the literature is the variety of methodological approaches, time periods, markets, and ESG measures used. In addition, Starks (2023) stresses that the concept of ESG can mean different things to different people in the sense that ESG can be seen through a *values* or a *value* lens. While the *values* view would most likely imply a negative association between ESG and stock returns (because investors are willing to trade off their ethical, societal, or religious values for lower returns), the *value* view would possibly imply a positive association between stock returns and ESG. For instance, some ESG metrics may allow investors to identify better governed companies (Gompers, Ishii, and Metrick 2003, Giroud and Mueller 2010, 2011) with more satisfied and possibly more productive employees (Edmans 2011).

Indeed, one puzzling aspect of our study is that it seems to contradict prior studies that have documented that certain subcomponents of ESG are correlated with stock returns. First, our study is not in contradiction with the idea that by cutting the data further and/or focusing on combinations of specific subperiods, countries, rating providers, or subcomponents of ESG, one might be able to tease out significant correlations. Our point is more that when studying the largest possible global sample of firms over the largest possible time period, using the most comprehensive set of ESG ratings up to date in a relatively agnostic and straightforward way, there is little evidence of a systematic relation between ESG ratings and stock returns.

However, we do not oppose the idea that more narrowly defined and more precisely measured subcomponents of ESG could give rise to correlations with stock returns. Indeed, the literature has shown that such correlations exist. For instance, Bolton and Kacperczyk (2021, 2023), Aswani, Raghunandan, and Rajgopal (2024), Zhang (2023) show that certain measures of GHG emissions (when appropriately lagged) can be correlated with stock returns. In a similar spirit, employee satisfaction Edmans (2011) also tends to correlate with stock returns, as do controversial business activities such as alcohol and tobacco (Hong and Kacperczyk 2009). We argue that these associations do not carry through to ESG ratings. Overall, it seems plausible and intuitive that studies focusing on ESG submeasures that are well-defined and more precisely measured could potentially give rise to more robust relations.

Finally, the fact that ESG ratings do not predict stock returns does not necessarily imply that ESG is uncorrelated with stock returns. Another possibility for the absence of a correlation between ESG ratings and stock returns is that ESG ratings do not accurately capture real corporate ESG actions and impacts, and investors agree with this. In a sense, prior evidence that real ESG actions that are clearly defined and measured such as reducing carbon footprints and improving employee welfare affect stock returns is consistent with this view. However, we do not think this reason to be a leading explanation for why ESG ratings are not correlated with stock returns, mainly because ESG ratings do capture meaningful ESG behavior. For instance, Derrien, Krueger, Landier, and Yao (2022) show that the occurrence of negative ESG incidents is inversely correlated with ESG ratings, which suggests that ESG ratings do capture, at least imperfectly, true ESG actions (or the lack thereof).

4 Conclusion

In contrast to prior studies, our analysis of a comprehensive global database (including 16,000+ stocks in 48 countries and seven different ESG rating providers over 2001-2020) uncovers very little evidence that ESG ratings are related to stock returns around the world. Drawing up the bill of 20 years of ESG investing, our results suggest that incorporating ESG ratings into investment strategies has not systematically come at the expense of financial returns. Our findings also suggest that the prices of stocks with high ESG ratings have not consistently been driven up by demand effects. On the flip side, we thus also do not find evidence of cost of capital effects of ESG ratings that could lead firms to internalize climate and social externalities (Fama 2021, Pástor et al. 2021). Further research is needed to examine whether the lack of a relation between ESG ratings and stock returns is due to the poor quality of ESG ratings, the less than pervasive prevalence of ESG preferences among investors, the challenges in distinguishing between short-term demand effects and long-term equilibrium effects, and/or other reasons.

References

Aghion P, Bénabou R, Martin R, Roulet A (2023) Environmental preferences and technological choices: Is market competition clean or dirty? *American Economic Review: Insights* 5(1):1–19.

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

Aswani J, Raghunandan A, Rajgopal S (2024) Are carbon emissions associated with stock returns? *Review of Finance* 28(1):75–106.

Atz U, Van Holt T, Liu ZZ, Bruno CC (2023) Does sustainability generate better financial performance? Review, meta-analysis, and propositions. *Journal of Sustainable Finance & Investment* 13(1):802–825.

Avramov D, Cheng S, Lioui A, Tarelli A (2022) Sustainable investing with ESG rating uncertainty. *Journal of Financial Economics* 145(2):642–664.

Bekaert G, Rothenberg R, Noguer M (2023) Sustainable investment – Exploring the linkage between alpha, ESG, and SDGs. *Sustainable Development* 31(5):3831–3842.

Benjamini Y, Krieger AM, Yekutieli D (2006) Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93(3):491–507.

Berg F, Koelbel JF, Rigobon R (2022) Aggregate confusion: The divergence of ESG ratings. *Review of Finance* 26(6):1315–1344.

Berg F, Lo AW, Rigobon R, Singh M, Zhang R (2023) Quantifying the returns of ESG investing: An empirical analysis with six ESG metrics, available at SSRN 4367367.

Berk JB, Van Binsbergen JH (2025) The impact of impact investing. *Journal of Financial Economics* 164:103972.

Bessembinder H, Chen TF, Choi G, Wei KJ (2019) Do global stocks outperform U.S. treasury bills?, available at SSRN 3415739.

Black B, Carvalho A, Érica Gorga (2012) What matters and for which firms for corporate governance in emerging markets? Evidence from Brazil (and other BRIK countries). *Journal of Corporate Finance* 18(4):934–952.

Bloom HS (1995) Minimum detectable effects: A simple way to report the statistical power of experimental designs. *Evaluation review* 19(5):547–556.

BNP Paribas (2021) The ESG global survey.

BNP Paribas (2023) The ESG global survey.

Bolton P, Kacperczyk M (2021) Do investors care about carbon risk? *Journal of Financial Economics* 142(2):517–549.

Bolton P, Kacperczyk M (2023) Global pricing of carbon-transition risk. *Journal of Finance* 78(6):3677–3754.

Botero J, Djankov S, Porta RL, de Silanes FL, Shleifer A (2004) The regulation of labor. *Quarterly Journal of Economics* 119(4):1339–1382.

Brandon RG, Krueger P, Schmidt PS (2021) ESG rating disagreement and stock returns. *Financial analysts journal* 77(4):104–127.

Campbell JY (1991) A variance decomposition for stock returns. *Economic Journal* 101(405):157–179.

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

Chaieb I, Langlois H, Scaillet O (2021) Factors and risk premia in individual international stock returns. *Journal of Financial Economics* 141(2):669–692.

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

Coles JL, Heath D, Ringgenberg MC (2022) On index investing. *Journal of Financial Economics* 145(3):665–683.

Dell'Erba M, Doronzo M (2023) Sustainability gatekeepers: ESG ratings and data providers. *University of Pennsylvania Journal of Business Law* 25(2):355–413.

Derrien F, Krueger P, Landier A, Yao T (2022) ESG news, future cash flows, and firm value. *Swiss Finance Institute Research Paper* (21-84).

Dyck A, Lins KV, Roth L, Wagner HF (2019) Do institutional investors drive corporate social responsibility? International evidence. *Journal of Financial Economics* 131(3):693–714.

Eccles RG, Stroehle JC (2018) Exploring social origins in the construction of ESG measures. Available at SSRN 3212685.

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

Edmans A (2023) The end of ESG. *Financial Management* 52(1):3–17.

Edmans A, Gosling T, Jenter D (2024a) Sustainable investing: Evidence from the field. *FEB-RN Research Paper* (18).

Edmans A, Pu D, Zhang C, Li L (2024b) Employee satisfaction, labor market flexibility, and stock returns around the world. *Management Science* 70(7):4357–4380.

Eskildsen M, Ibert M, Jensen TI, Pedersen LH (2024) In search of the true greenium. *Available at SSRN* .

European Commission (2023) Proposal for a regulation of the european parliament and of the council on the transparency and integrity of Environmental, Social and Governance (ESG).

Fama E (2021) Contract costs, stakeholder capitalism, and ESG. *European Financial Management* 27(2):189–195.

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

Fama E, French K (2015) A five-factor asset pricing model. *Journal of Financial Economics* 116(1):1–22.

Fama E, MacBeth J (1973) Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81(3):607–636.

Fitzgibbons S, Pedersen LH, Pomorski L (2021) Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics* 142(2):572–597.

Friede G, Busch T, Bassen A (2015) ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment* 5(4):210–233.

FT (2017) The ethical investment boom. *Financial Times* URL `https://www.ft.com/content/9254dfd2-8e4e-11e7-a352-e46f43c5825d`.

Geczy CC, Stambaugh RF, Levin D (2021) Investing in socially responsible mutual funds. *The Review of Asset Pricing Studies* 11(2):309–351.

Gibson R, Krueger P, Riand N, Schmidt P (2021) ESG rating disagreement and stock returns. *Financial Analysts Journal* 77(4):104–127.

Gillan S, Koch A, Starks L (2021) Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance* 66, 101889.

Giroud X, Mueller HM (2010) Does corporate governance matter in competitive industries? *Journal of Financial Economics* 95(3):312–331.

Giroud X, Mueller HM (2011) Corporate governance, product market competition, and equity prices. *Journal of Finance* 66(2):563–600.

Global Sustainable Investment Alliance (GSIA) (2021) Global sustainable investment review.

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

Green J, Hand J, Zhang XF (2020) The characteristics that provide independent information about average U.S. monthly stock returns. *Review of Financial Studies* 30(12):4389–4436.

Hamilton S, Jo H, Statman M (1993) Doing well while doing good? The investment performance of socially responsible mutual funds. *Financial analysts journal* 49(6):62–66.

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

Harvey CR, Liu Y, Zhu H (2016) . . . and the cross-section of expected returns. *Review of Financial Studies* 29(1):5–68.

Heath D, Macciocchi D, Michaely R, Ringgenberg MC (2023a) Does socially responsible investing change firm behavior? *Review of Finance* 27(6):2057–2083.

Heath D, Ringgenberg MC, Samadi M, Werner IM (2023b) Reusing natural experiments. *Journal of Finance* 78(4):2329–2364.

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

Hou K, Hsu PH, Wang S, Watanabe A, Xu Y (2022) Corporate R&D and stock returns: International evidence. *Journal of Financial and Quantitative Analysis* 57(4):1377–1408.

Hou K, Kho BC, Karolyi A (2011) What factors drive global stock returns? *Review of Financial Studies* 24(8):2527–2574.

Ilhan E, Krueger P, Sautner Z, Starks LT (2023) Climate risk disclosure and institutional investors. *Review of Financial Studies* 36(7):2617–2650.

International Organization of Securities Commissions (IOSCO) (2021) Environmental, Social, and Governance (ESG) ratings and data product providers.

Karolyi GA, Wu Y, Xiong WW (2023) Understanding the global equity greenium, available at SSRN 4391189.

Kempf A, Osthoff P (2007) The effect of socially responsible investing on financial performance. *European Financial Management* 13:908–922.

Klausmann J, Krueger P, Matos P (2024) The green transition: Evidence from corporate green revenues. *Available at SSRN 4850449* .

Krueger P (2015) Corporate goodness and shareholder wealth. *Journal of Financial Economics* 115:304–329.

Krueger P, Sautner Z, Tang DY, Zhong R (2024) The effects of mandatory ESG disclosure around the world. *Journal of Accounting Research* 62(5):1795–1847.

Liang H, Renneboog L (2021) Corporate social responsiblity and sustainable finance: A review of the literature. *Oxford Research Encyclopedia of Economics and Finance* .

Lindsey LA, Pruitt S, Schiller C (2023) The cost of ESG investing., available at SSRN 3975077.

Martínez-Zarzoso ABM Inmaculada, Morales-Lage R (2019) Does environmental policy stringency foster innovation and productivity in OECD countries? *Energy Policy* 134, 110982.

Newey WK, West KD (1987) Hypothesis testing with efficient method of moments estimation. *International Economic Review* 777–787.

Pástor L, Stambaugh R, Taylor L (2021) Sustainable investing in equilibrium. *Journal of Financial Economics* 142(2):550–571.

Pástor L, Stambaugh R, Taylor L (2022) Dissecting green returns. *Journal of Financial Economics* 146(2):403–424.

Renneboog L, Ter Horst J, Zhang C (2008) Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of banking & finance* 32(9):1723–1742.

Shanaev S, Ghimire B (2022) When ESG meets AAA: The effect of ESG rating changes on stock returns. *Finance Research Letters* 46:102302.

Simeth M, Wehrheim D (2024) Innovation and institutional ownership. *Journal of Corporate Finance* 86:102569.

Starks LT (2023) Presidential address: Sustainable finance and ESG issues—value versus values. *Journal of Finance* 78(4):1837–1872.

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

SustainAbility (2020) Rate the raters 2020: Investor survey and interview results, SustainAbility Rate the Raters Research Series.

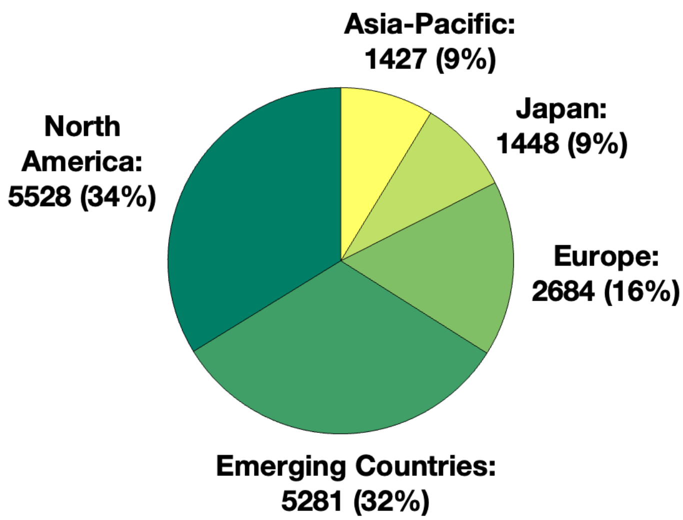
Van der Beck P (2024) Flow-driven ESG returns. *Swiss Finance Institute Research Paper* (21-71).

Welzel C (2013) *Freedom Rising: Human Empowerment and the Quest for Emancipation* (United Kingdom. Cambridge University).

Zhang S (2023) Carbon returns across the globe. *Journal of Finance* Forthcoming.

Figure 1. Sample coverage by region and country, January 2001 - December 2020

The pie chart on top shows the number and percentage of unique stocks in our sample that are publicly traded in each geographic region between January 2001 and December 2020. The bottom pie chart shows the number and percentage of sample stocks that are publicly traded in each country during the same period. To ensure readability, only the ten countries with the greatest number of stocks are shown. The category Other Countries refers to the remaining 38 countries in the sample.



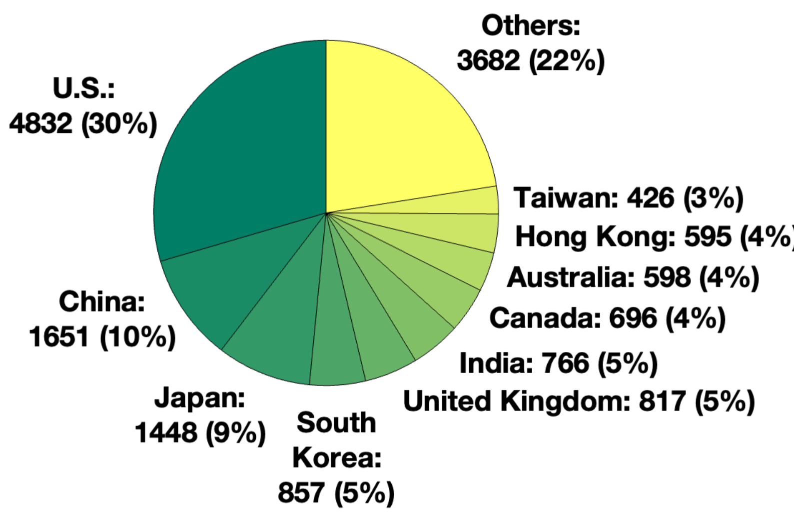


Figure 2. Correlations of ESG ratings across rating agencies

This figure summarizes the full sample correlations of ESG ratings across rating agencies for each geographic region and for the world. The first column reports the average correlations across all rating-pairs. The table also shows the minimum and maximum correlation across rating-pairs. Since the RepRisk rating is negatively correlated with the other ratings, we also report averages and minima of correlations across all rating-pairs other than RepRisk. The correlations are based on all firms in our sample and on the full sample period from January 2001 until December 2020.

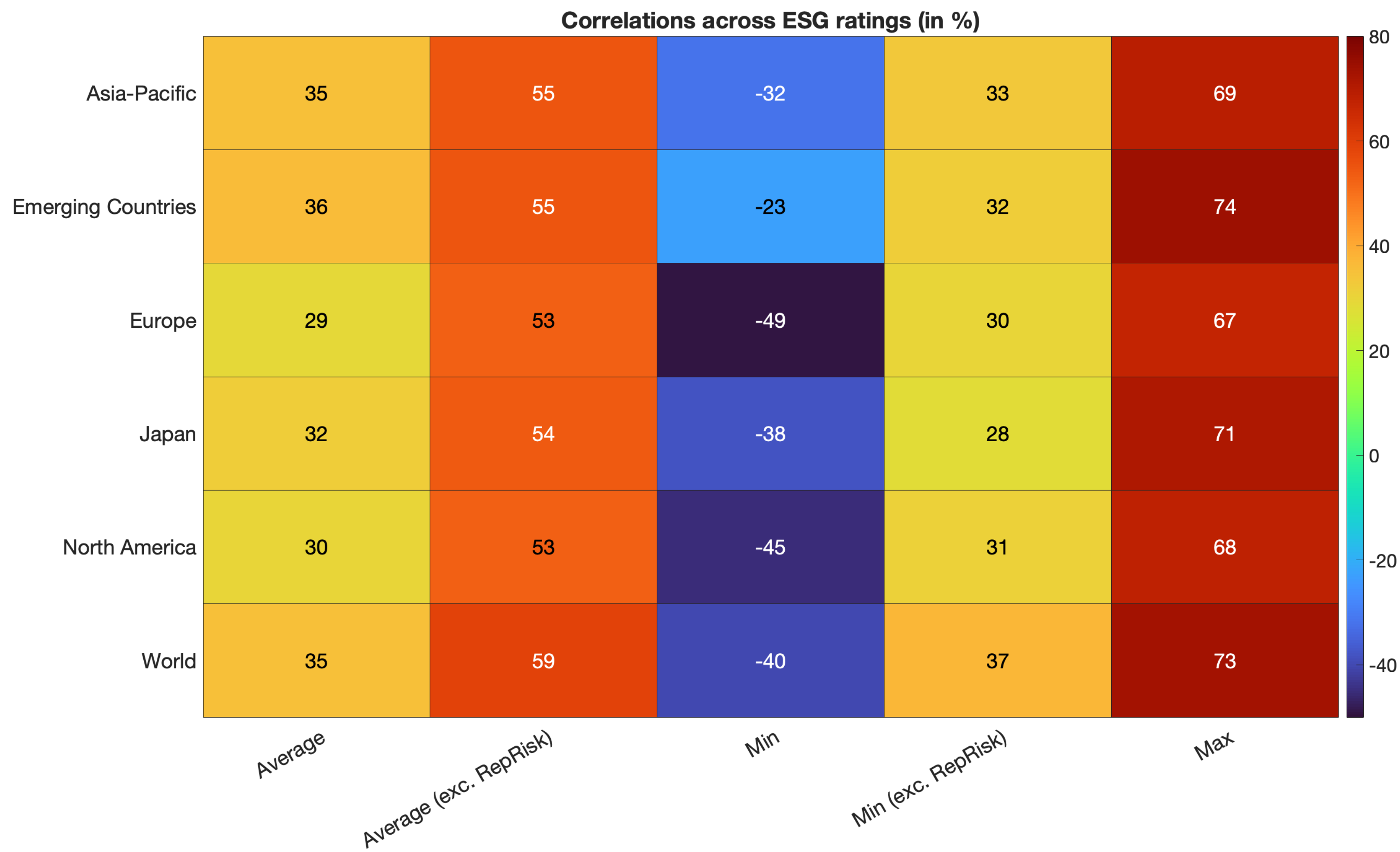
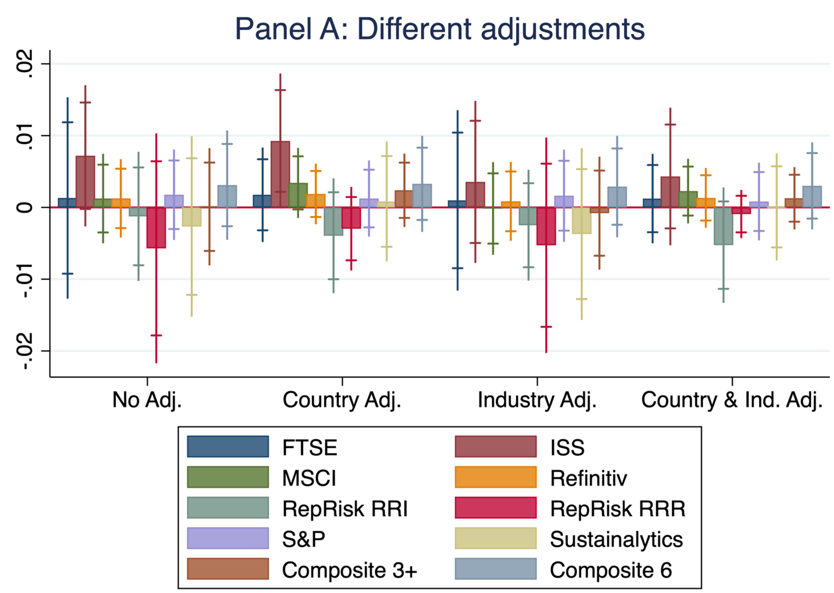


Figure 3. ESG ratings and stock returns: Variations of the baseline results

This figure shows the results of several variations of the baseline Fama-Macbeth regressions in Table 1 of monthly stock returns on lagged ESG ratings using the following ten different ratings: FTSE, ISS, MSCI IVA, Refinitiv, RepRisk (RRI), RepRisk (RRR), S&P Global, Sustainalytics, Composite 3+, and Composite 6. Panels A-J present the results of ten different variations of the baseline regressions. Each panel presents, in bars, the coefficient on the different ESG ratings obtained in separate regressions that include the full set of control variables from Table 1. Confidence intervals are depicted as whiskers around the point estimates. The longer (shorter) whiskers represent 99% (95%) confidence intervals. Confidence intervals use Newey and West (1987) standard errors with automatic lag selection. Panel A shows the results when using country- and/or industry-adjusted ESG ratings, where the baseline results based on unadjusted ratings are included at the left-hand side of the panel. Panel B shows the results for the E, S, and G subratings. Panel C shows the results for different geographic regions. Panel D shows the results for the the period after the Paris Agreement (January 2016 to December 2020), using unadjusted, country- and/or industry-adjusted ESG ratings. Panel E shows the results for the E subrating separately for the U.S. and the rest of the world excluding U.S. (RoW) controlling either for the full set of controls or for the firm characteristics that correspond to the factors of the Fama and French (2015) five-factor model plus momentum (FF6). Panel F reports the results for ESG momentum, defined as the most recent change in ESG ratings during the previous 12 months, using unadjusted, country- and/or industry-adjusted ESG ratings. Panels G and H show the results for ESG downgrades and upgrades, respectively, under different adjustments and using the global sample. Panels I (Panel J) reports the results for worst-in-class and best-in-class ESG strategies using the global sample (U.S. subsample). The sample period is from January 2001 to December 2020, but the starting dates vary across raters and regions.



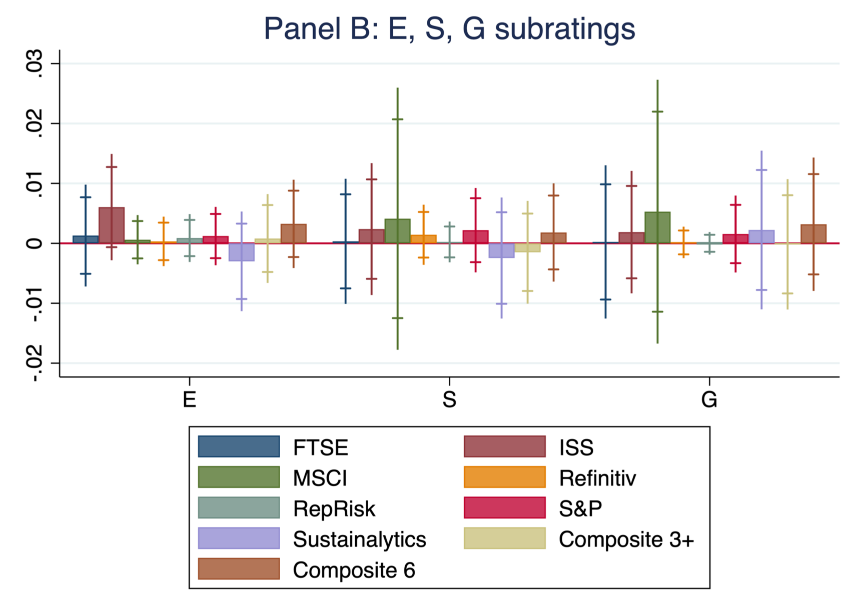


Figure 3 - continued

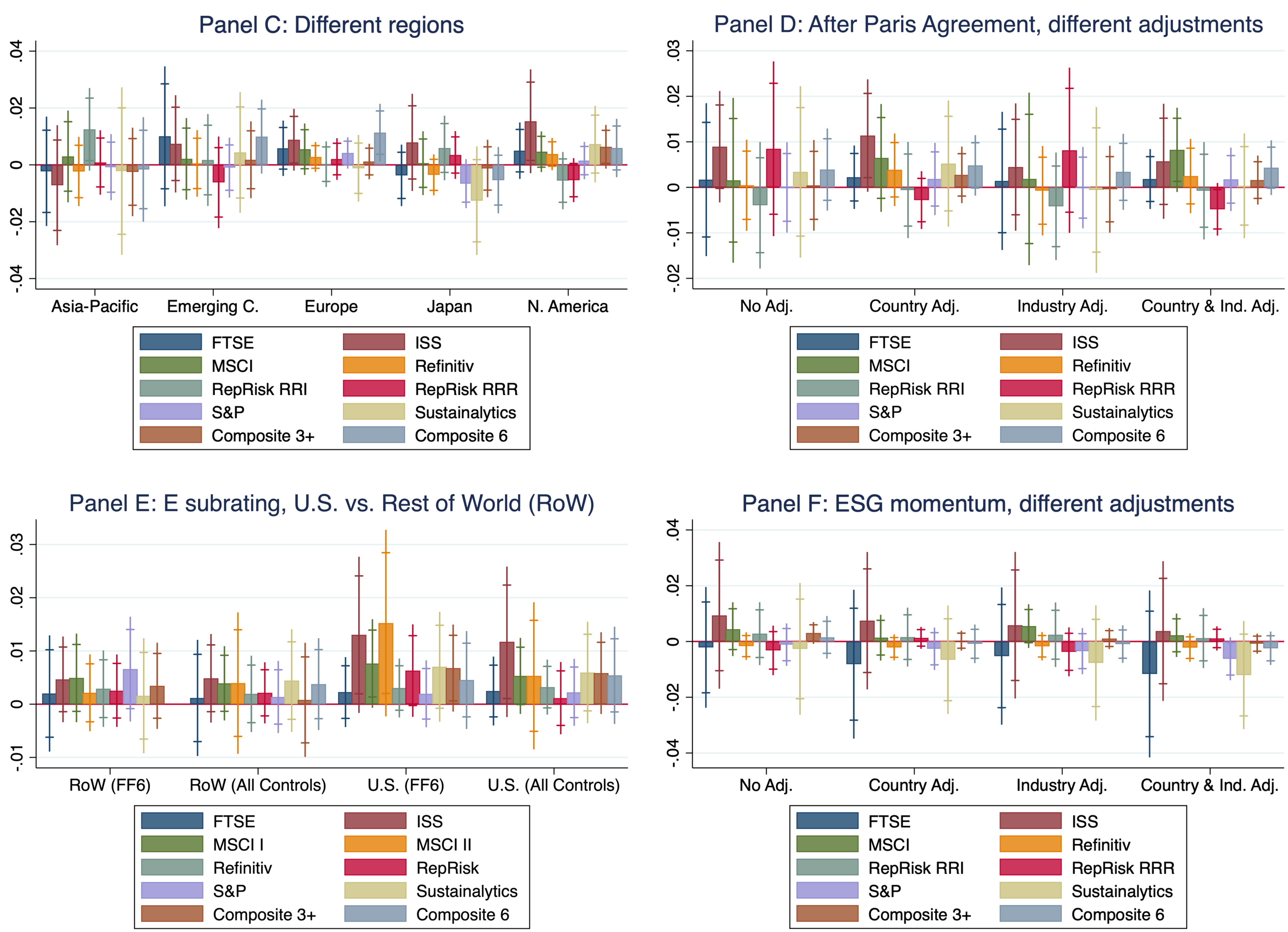


Figure 3 - continued

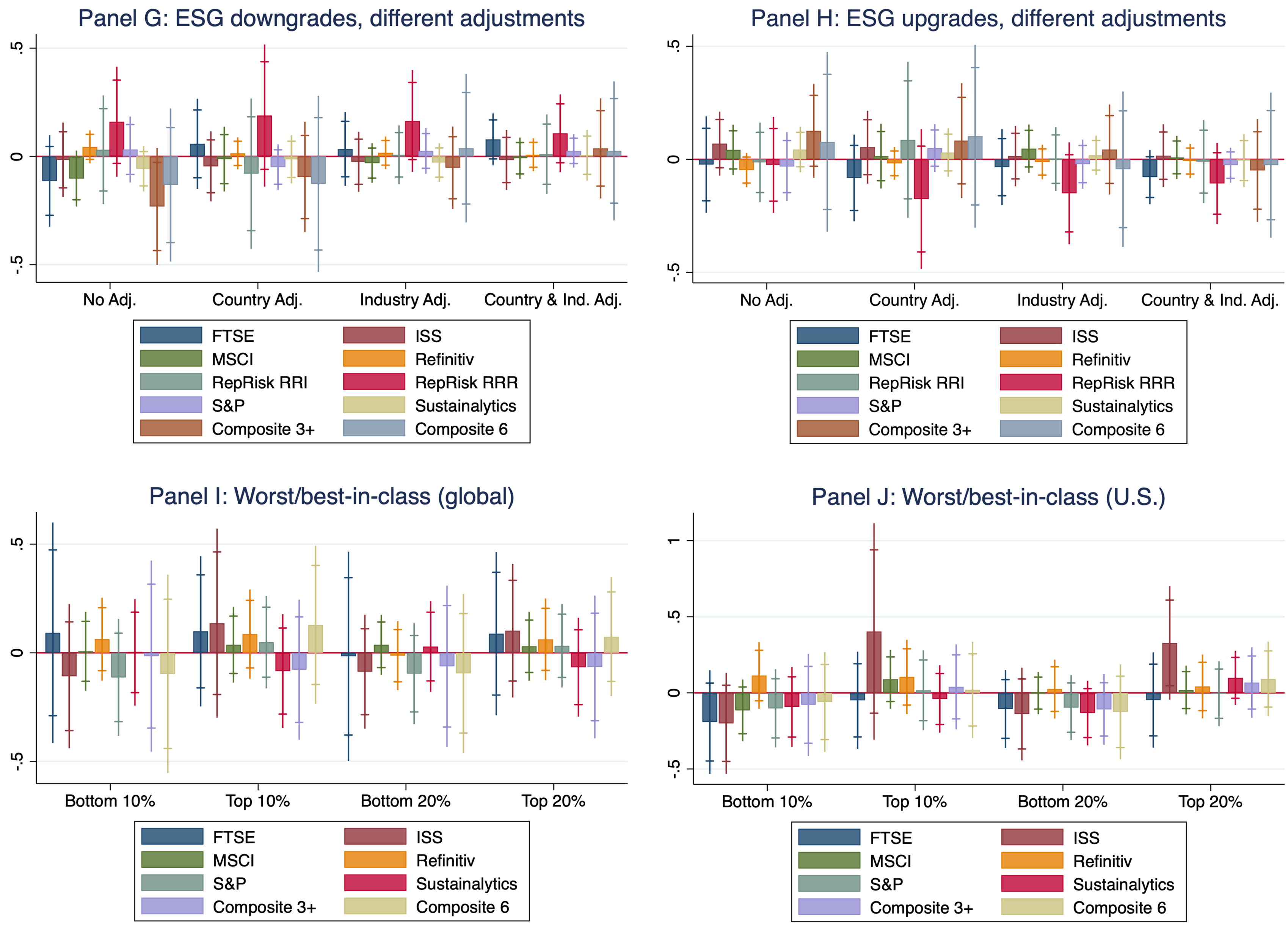


Table 1. Baseline results on the relation between ESG ratings and stock returns

This table reports the results of Fama-Macbeth regressions of monthly stock returns on lagged ESG ratings using the global sample of stocks. Each column (1) through (10) shows the results for a different ESG rating: FTSE, ISS, MSCI IVA, Refinitiv, RepRisk (RRI), RepRisk (RRR), S&P Global, Sustainalytics, Composite 3+, and Composite 6. The table also presents the 95% confidence interval for each ESG rating. Variable descriptions are in the Internet Appendix. The sample period is from January 2001 to December 2020, but the starting dates vary across ratings as indicated in bottom row of the table. t-statistics based on Newey and West (1987) standard errors with automatic lag selection are reported in parentheses. ***, ** and * denote statistical significance at, respectively, the 1%, 5% and 10% levels.

	FTSE (1)	ISS (2)	MSCI (3)	Refinitiv (4)	RRI (5)	RRR (6)	S&P (7)	Sustainalytics (8)	Composite 3+ (9)	Composite 6 (10)
ESG	0.001	0.007*	0.001	0.001	-0.001	-0.006	0.002	-0.003	0.000	0.003
	(0.249)	(1.930)	(0.515)	(0.597)	(-0.363)	(-0.929)	(0.728)	(-0.555)	(0.029)	(1.085)
Beta	0.124	0.274	-0.116	0.055	0.021	0.049	0.049	0.081	0.190	0.197
	(0.652)	(1.578)	(-0.891)	(0.459)	(0.176)	(0.440)	(0.428)	(0.498)	(1.088)	(0.863)
Size	-0.028	0.004	-0.088**	-0.113**	-0.150***	-0.158***	-0.039	-0.015	0.028	-0.054
	(-0.395)	(0.088)	(-2.284)	(-2.338)	(-3.498)	(-3.544)	(-0.909)	(-0.336)	(0.438)	(-0.755)
B/M	-0.090*	-0.029	0.050	0.024	-0.036	-0.028	-0.003	-0.014	-0.059	-0.101**
	(-1.708)	(-0.545)	(0.921)	(0.541)	(-0.596)	(-0.499)	(-0.077)	(-0.288)	(-1.205)	(-2.304)
Investment	-0.042	-0.198	-0.123	-0.081	-0.205**	-0.195**	-0.108	-0.057	0.021	-0.003
	(-0.406)	(-1.331)	(-1.497)	(-0.944)	(-2.432)	(-2.395)	(-1.058)	(-0.586)	(0.203)	(-0.026)
Profitability	0.230	0.508*	0.389***	0.504***	0.343	0.352	0.436**	0.605***	0.396	0.174
	(0.858)	(1.895)	(2.628)	(2.798)	(1.371)	(1.507)	(2.396)	(2.793)	(1.467)	(0.641)
Momentum	0.004	0.007**	0.001	0.002	0.001	0.001	0.003	0.007**	0.005	0.004
	(0.893)	(2.163)	(0.130)	(0.541)	(0.186)	(0.151)	(0.753)	(2.295)	(1.249)	(0.788)
Volatility	0.006	-0.004	-0.000	-0.001	-0.004	-0.005	-0.005	-0.006	-0.003	0.010
	(0.593)	(-0.464)	(-0.084)	(-0.243)	(-0.776)	(-1.116)	(-0.740)	(-0.775)	(-0.381)	(0.861)
Leverage	-0.438	-0.085	-0.142	-0.096	-0.351	-0.368	-0.086	0.027	-0.236	-0.465**
	(-1.575)	(-0.326)	(-0.612)	(-0.371)	(-1.293)	(-1.382)	(-0.343)	(0.097)	(-0.859)	(-2.027)
Tangibility	-0.112	-0.556*	-0.095	-0.011	0.059	0.003	0.051	-0.419	-0.248	-0.057
	(-0.376)	(-1.712)	(-0.394)	(-0.044)	(0.215)	(0.011)	(0.204)	(-1.368)	(-0.724)	(-0.179)
R&D	5.960***	4.138***	2.730**	2.049*	3.652***	3.947***	3.167**	3.795***	5.270***	5.727**
	(2.907)	(2.829)	(2.451)	(1.923)	(2.901)	(3.470)	(2.441)	(2.807)	(3.308)	(2.324)
95% CI	[-0.009;0.012]	[0.000;0.015]	[-0.003;0.006]	[-0.003;0.005]	[-0.008;0.006]	[-0.018;0.006]	[-0.003;0.007]	[-0.012;0.007]	[-0.006;0.006]	[-0.003;0.009]
R^2	0.081	0.063	0.092	0.085	0.048	0.056	0.086	0.082	0.071	0.093
No. obs.	188,509	261,490	668,880	507,195	1,208,160	1,208,160	395,688	380,497	282,124	116,404
Start date	2015-Jan	2013-Apr	2001-Jan	2003-Jul	2007-Feb	2007-Feb	2003-Jul	2009-Sep	2014-Jan	2015-Jan

Table 2. Value-weighted portfolio sorts adjusted for multiple hypothesis testing

This table summarizes the alphas of value-weighted portfolio strategies based on ESG ratings. We sort stocks at the end of each month into high and low ESG rating portfolios and compute the return spread (high minus low) in excess of the one-month U.S. Treasury bill rate. Stocks are value-weighted based on their market capitalization at the end of the previous month. The sorts are conducted separately for all combinations of geographic regions, ESG (sub)ratings, country and/or industry rating adjustments, and three alternative portfolio breakpoints (top and bottom 10%, 20%, or 30%). The table summarizes raw return spreads as well as alphas relative to five different popular factor models: the CAPM, the Fama and French (1993) three-factor model (*FF3*), the Carhart (1997) four-factor model (*Carhart*), the Fama and French (2015) five-factor model (*FF5*), and the five-factor model plus momentum (*FF5+MOM*)). For each combination of (sub)rating, risk model, geographic region, and rating adjustment, the table reports two numbers: x/y. Here, x denotes the number of tests where the time-series alpha is statistically significant at the 5% level or better, and y represents the number of those significant alphas that are positive. For example, a hypothetical entry "6/5" indicates that, out of all tests for that specific combination, six alphas are statistically significant, and five of those are positive. The total number of tests per cell/combination is indicated at the bottom of the table, either 27 or 24, depending on whether ESG ratings or subratings are used. Specifically, 27 (24) tests are performed when using nine (eight) ESG ratings (subratings) and three alternative breakpoints. We use the following eight ESG subrating sources to form portfolios: FTSE, ISS, MSCI IVA, Refinitiv, S&P Global, Sustainalytics, Composite 3+, and Composite 6. Additionally, the RRR RepRisk rating is used for portfolios formed on ESG ratings. We exclude the RRI/E/S/G RepRisk ratings, as they exhibit limited cross-sectional variation across firms for sorting purposes. Significance tests are corrected for multiple hypothesis testing using the procedure of Benjamini et al. (2006), applied per column (i.e., separately for each (sub)rating and each rating adjustment). The sample period is from January 2001 to December 2020, but the exact starting date for each region is chosen so that there are at least 120 stocks in that region as of the starting date. Variable descriptions are provided in the Internet Appendix.

	No Adjustment				Industry-adjusted				Country-adjusted				Country & ind. adjusted			
	ESG (1)	E (2)	S (3)	G (4)	ESG (5)	E (6)	S (7)	G (8)	ESG (9)	E (10)	S (11)	G (12)	ESG (13)	E (14)	S (15)	G (16)
Asia-Pacific																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0	3/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	4/0	0/0	0/0	0/0	0/0	0/0	0/0
Emerging Countries																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	4/4	1/1	1/1	0/0	2/2	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	4/4	1/1	1/1	0/0	3/3	0/0	0/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	2/2	1/1	1/1	0/0	2/2	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/1	0/0	0/0	1/1	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0

Table 2 - continued

	No Adjustment				Industry-adjusted				Country-adjusted				Country & ind. adjusted			
	ESG (1)	E (2)	S (3)	G (4)	ESG (5)	E (6)	S (7)	G (8)	ESG (9)	E (10)	S (11)	G (12)	ESG (13)	E (14)	S (15)	G (16)
Europe																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	2/0	0/0	0/0	1/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0	1/0	0/0	0/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Japan																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
North America																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	2/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	2/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0
No. raters	9	8	8	8	9	8	8	8	9	8	8	8	9	8	8	8
No. hypotheses per cell	27	24	24	24	27	24	24	24	27	24	24	24	27	24	24	24
No. hypotheses per column	810	720	720	720	810	720	720	720	810	720	720	720	810	720	720	720

Table 3. Effect sizes and confidence intervals

This table summarizes the effect sizes and 95% confidence intervals (CIs) corresponding to the baseline Fama-Macbeth regression results presented in Figure 3. For each panel in Figure 3, we report the mean, median, and the 25^{th} and 75^{th} percentiles of the effect size distribution for the estimates shown in that panel, along with the median of the lower and upper bounds of the corresponding 95% confidence intervals. Both effect sizes and confidence intervals are expressed in annualized percentage returns in response to a one-sample standard deviation shock to a given ESG predictor.

	Annualized effect sizes (in %)				Annualized 95% CIs (in %)	
	Mean (1)	25^{th} Percentile (2)	Median (3)	75^{th} Percentile (4)	Median LB (5)	Median UB (6)
Panel A – different adjustments	0.175	-0.16	0.25	0.459	-0.678	1.201
Panel B – E/S/G subratings	0.31	0.063	0.273	0.465	-0.973	1.697
Panel C – different regions	0.334	-0.437	0.305	1.005	-1.178	1.803
Panel D – after Paris Agreem., diff. adj.	0.451	-0.01	0.416	0.775	-1.036	1.852
Panel E – E subrating, U.S. vs. RoW	1.048	0.659	0.958	1.246	-0.581	2.515
Panel F – ESG momentum, diff. adj.	-0.068	-0.206	-0.123	0.134	-0.541	0.488
Panel G – ESG downgrades, diff. adj.	-0.038	-0.568	0.077	0.395	-1.326	1.217
Panel H – ESG upgrades, diff. adj.	-0.018	-0.345	-0.078	0.521	-1.267	1.434
Panel I – best-in-class (global)	0.472	-0.211	0.664	1.121	-1.664	2.487
Panel I – worst-in-class (global)	-0.355	-1.144	-0.189	0.213	-3.447	2.205
Panel J – best-in-class (U.S.)	0.883	0.069	0.49	1.139	-1.501	2.815
Panel J – worst-in-class (U.S.)	-1.057	-1.546	-1.246	-0.831	-3.538	1.103

Table 4. Minimum detectable effect sizes (MDES)

This table summarizes the minimum detectable effect sizes (MDES) of Bloom (1995) corresponding to the baseline Fama-Macbeth regression results presented in Figure 3. For each panel in Figure 3, we report the mean, median, and the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles of the distribution of MDES for the estimates displayed in that panel. In Panel A, the MDES is expressed as a fraction of the sample standard deviation of the dependent variable (monthly stock returns). For instance, a hypothetical value of "1/30" indicates that we can reliably detect effects smaller than 1/30$^{\text{th}}$ of one standard deviation of the dependent variable. In Panel B, we report the MDES in annualized percentage returns in response to a one-sample standard deviation shock to a given ESG predictor. For instance, a hypothetical MDES of ±1% implies that a one-sample standard deviation shock to the ESG predictor, resulting in a change in annualized monthly stock returns of 1 percentage point (annualized) or less, can be reliably detected.

Panel A: MDES relative to one sample standard deviation of monthly stock returns

	Mean (1)	25$^{\text{th}}$ Percentile (2)	Median (3)	75$^{\text{th}}$ Percentile (4)
Panel A – different adjustments	1/110	1/144	1/100	1/70
Panel B – E/S/G subratings	1/82	1/93	1/72	1/50
Panel C – different regions	1/65	1/83	1/57	1/43
Panel D – after Paris Agreem., diff. adj.	1/80	1/93	1/73	1/58
Panel E – E subrating, U.S. vs. RoW	1/63	1/74	1/62	1/52
Panel F – ESG momentum, diff. adj.	1/208	1/287	1/197	1/143
Panel G – ESG downgrades, diff. adj.	1/76	1/93	1/73	1/47
Panel H – ESG upgrades, diff. adj.	1/78	1/97	1/76	1/51
Panel I – best/worst-in-class (global)	1/41	1/53	1/37	1/27
Panel J – best/worst-in-class (U.S.)	1/44	1/53	1/42	1/35

Panel B: MDES expressed as annualized returns (in %)

	Mean (1)	25$^{\text{th}}$ Percentile (2)	Median (3)	75$^{\text{th}}$ Percentile (4)
Panel A – different adjustments.	± 1.601	± 1.041	± 1.405	± 1.846
Panel B – E/S/G subratings	± 2.157	± 1.504	± 2.021	± 2.640
Panel C – different regions	± 2.507	± 1.573	± 2.151	± 3.015
Panel D – after Paris Agreem., diff. adj.	± 2.130	± 1.537	± 1.914	± 2.466
Panel E – E subrating, U.S. vs. RoW	± 2.323	± 1.901	± 2.317	± 2.581
Panel F – ESG momentum, diff. adj.	± 0.865	± 0.544	± 0.658	± 0.992
Panel G – ESG downgrades, diff. adj.	± 2.311	± 1.361	± 2.013	± 3.132
Panel H – ESG upgrades, diff. adj.	± 2.237	± 1.379	± 1.909	± 2.713
Panel I – best/worst-in-class (global)	± 3.672	± 2.566	± 3.412	± 4.612
Panel J – best/worst-in-class (U.S.)	± 3.434	± 2.741	± 3.193	± 3.973

Table 5. Is the ESG-return relation stronger when there is less ESG uncertainty?

This table reports the results of Fama-Macbeth regressions of monthly stock returns on lagged ESG ratings, a dummy variable indicating Low ESG uncertainty, and the interaction of lagged ESG ratings with the Low ESG uncertainty dummy. In this table the ESG variable is defined as the average ESG rating across all six ESG raters used throughout the paper other than RepRisk. ESG ratings are converted to percentile ranks in each month before averaging. ESG uncertainty is the standard deviation of these percentile ranks across these six raters. Low ESG Uncertainty is a dummy variable equal to one if a stock is among the 20% of stocks with lowest ESG uncertainty in a given month. ESG and Low ESG uncertainty are missing if fewer than two raters rate a stock in a given month. Panels A, B, and C show the results for the following three sample periods: 2003-2013, 2014-2020, and 2003-2020, respectively. Columns (1) through (7) in each panel show results by region (Asia-Pacific, Emerging Countries, Europe, Japan, North America), pooled across regions (Global), as well as U.S. The exact starting date for each region is chosen so that there are at least 120 stocks in that region as of the starting date. All regressions include the full set of control variables listed in Table 1, but coefficients are suppressed for space considerations. Variable descriptions are in the Internet Appendix. t-statistics based on Newey and West (1987) standard errors with automatic lag selection are reported in parentheses. ***, ** and * denote statistical significance at, respectively, the 1%, 5% and 10% levels.

Panel A: ESG Uncertainty, 2003-2013

	Asia-Pacific	Emerging Countries	Europe	Japan	North America	Global	U.S.
Low ESG uncertainty	-0.502	0.081	-0.097	0.181	0.247	0.102	0.260
	(-1.312)	(0.266)	(-0.488)	(0.614)	(1.512)	(0.929)	(1.627)
ESG	-0.007	-0.006	0.002	-0.004	0.003	0.002	0.005*
	(-1.099)	(-0.977)	(0.939)	(-1.004)	(1.501)	(1.068)	(1.770)
Interaction	0.015*	0.001	-0.001	-0.003	-0.006**	-0.002	-0.007**
	(1.835)	(0.227)	(-0.270)	(-0.565)	(-2.052)	(-0.902)	(-2.146)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.196	0.138	0.131	0.191	0.141	0.098	0.145
No. obs.	15,216	13,186	53,859	26,147	81,240	197,789	70,993

Table 5 - continued

Panel B: ESG Uncertainty, 2014-2020							
	Asia-Pacific	Emerging Countries	Europe	Japan	North America	Global	U.S.
Low ESG uncertainty	0.001	0.050	-0.160	-0.308**	0.223	0.053	0.289
	(0.004)	(0.318)	(-0.595)	(-2.000)	(1.300)	(0.577)	(1.633)
ESG	-0.003	0.001	0.001	-0.004	0.009***	0.001	0.009***
	(-0.434)	(0.288)	(0.383)	(-0.945)	(2.907)	(0.239)	(2.922)
Interaction	-0.006	0.003	0.004	0.007**	-0.002	0.000	-0.004
	(-1.073)	(0.980)	(1.329)	(2.313)	(-0.877)	(0.200)	(-1.333)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.117	0.083	0.079	0.129	0.119	0.066	0.116
No. obs.	33,593	70,085	80,584	40,333	143,437	368,032	127,488

Panel C: ESG Uncertainty, 2003-2020							
	Asia-Pacific	Emerging Countries	Europe	Japan	North America	Global	U.S.
Low ESG uncertainty	-0.220	0.058	-0.123	-0.040	0.237**	0.083	0.272**
	(-1.012)	(0.413)	(-0.787)	(-0.232)	(1.984)	(1.133)	(2.300)
ESG	-0.005	-0.000	0.002	-0.004	0.006***	0.001	0.006***
	(-1.025)	(-0.117)	(0.972)	(-1.353)	(3.060)	(0.905)	(3.216)
Interaction	0.003	0.002	0.001	0.002	-0.005**	-0.001	-0.006**
	(0.592)	(0.952)	(0.533)	(0.526)	(-2.151)	(-0.747)	(-2.510)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.152	0.098	0.110	0.163	0.132	0.085	0.133
No. obs.	48,809	83,271	134,443	66,480	224,677	565,821	198,481

Table 6. Does the ESG-return relation depend on country characteristics?

This table reports the results of panel regressions of monthly stock returns on lagged ESG (sub)ratings and the interaction of lagged ESG (sub)ratings with various country characteristics. All regressions control for the full set of stock characteristics used in Table 1 as well as country-month fixed effects. The country characteristics are classified in three broad categories: ESG social norms (Panel A), ESG disclosure standards (Panel B), and ESG regulations (Panel C). For each (sub)rating and each country characteristic, the table reports two numbers x/y, where x denotes the number of ratings (out of the ten different ratings from Table 1) for which the interaction effect between that (sub)rating and country characteristic is statistically significant at the 5% level or better, and y denotes the number of statistically significant interaction effects that have a positive sign. For example, a hypothetical entry "5/3" in the table would mean that, out of the ten different ESG ratings, there are five statistically significant interaction effects with a given country characteristic, and that three of those are positive. For the ESG rating in column (1) of each panel, we use ten different ratings and test for ten different interaction effects for each country characteristic. For the E, S, and G subratings in columns (2)-(4) of each panel, we use nine different ratings since RepRisk only reports one set of subratings. Statistical significance is based on standard errors double clustered at the stock and month levels. The sample period is from January 2014 to December 2020. Variable descriptions are in the Internet Appendix.

Panel A: ESG social norms

	ESG (1)	E (2)	S (3)	G (4)	# Countries (5)
ESG attitudes & beliefs					
Environmental norms	0/0	0/0	0/0	0/0	46
Social norms	0/0	0/0	0/0	0/0	46
Associational act.	1/1	0/0	0/0	1/1	38
Social movement act.	0/0	0/0	0/0	1/0	45
Political orientation					
Schmidt index	0/0	0/0	0/0	0/0	22
Votes green	0/0	1/0	0/0	1/1	22
Votes left I	1/1	1/1	1/1	1/1	22
Votes left II	0/0	0/0	0/0	0/0	22
Votes left III	0/0	0/0	0/0	1/0	22
Votes non-right I	1/1	0/0	0/0	1/1	22
Votes non-right II	0/0	0/0	0/0	0/0	22
Votes non-right III	1/1	0/0	0/0	1/0	22
Left-wing gov. I	0/0	0/0	0/0	0/0	22
Left-wing gov. II	0/0	0/0	0/0	0/0	22
Left-wing gov. III	0/0	0/0	0/0	0/0	22

Panel B: ESG disclosure standards

	ESG (1)	E (2)	S (3)	G (4)	# Countries (5)
ESG disclosure standards					
Mand. ESG	0/0	0/0	0/0	0/0	48
Mand. E	1/1	1/1			48
Mand. S	0/0		0/0		48
Mand. G	1/1			0/0	48
Mand. ESG (all at once)	0/0	0/0	1/1	1/1	48
Mand. ESG (gov. mand.)	1/1	1/1	1/1	0/0	48
Mand. ESG (full comp.)	0/0	0/0	0/0	0/0	48

Panel C: ESG regulations

	ESG (1)	E (2)	S (3)	G (4)	# Countries (5)
Strictness of ESG regulations					
Env. Performance	0/0	0/0		1/0	47
Env. Democracy	0/0	0/0		0/0	25
Env. Policy Stringency	1/0	1/0		1/0	25
Employment Laws index	1/1		1/1	0/0	46
Labor Regulation index	0/0		0/0	0/0	48
Employment Protection I	0/0		0/0	0/0	28
Employment Protection II	0/0		0/0	0/0	28

Internet Appendix to "Drawing Up the Bill: Is ESG Related to Stock Returns Around the World?"

A Data Appendix

In this appendix, we detail how we construct and clean the global stock returns dataset. We base our filters on both Bessembinder et al. (2019) and Chaieb et al. (2021). To our knowledge, these papers use the most thorough sets of filters in the literature.

A.1 Stock return data

We collect data for U.S. and Canadian stocks from CRSP and Compustat North America, respectively. Data for the remaining stocks are sourced from Compustat Global. We acknowledge that many papers use Datastream as a source of data for international stock returns. We use Compustat Global instead. Our main motivation is that Chaieb et al. (2021) conduct an in-depth comparison of both databases and conclude that Compustat Global has considerably fewer errors than Datastream. Compustat Global also differs from Datastream in that it distinguishes between types of daily quotes (e.g., the difference between a closing price and a price that is carried forward).

For U.S. stocks, we use the provided monthly level returns (CRSP field *RET*), the absolute value of end-of-month closing prices (CRSP field *ALTPRC*), the number of shares outstanding (CRSP field *SHROUT*), the stock split adjustment factor for shares outstanding (CRSP field *CFACSHR*), and the stock split adjustment factor for prices (CRSP field *CFAPCR*). We follow Bali, Engle, and Murray (2016) in the use of *ALTPRC* instead of *PRC*. For the exact formulas used to compute all variables used in the paper please refer to Appendix Table IA.1.

For non-U.S. stocks, we follow Bessembinder et al. (2019) and compute stock returns and market capitalization from Compustat Global daily data on closing prices (Compustat field *PRCCD*), number of shares outstanding (Compustat field *CSHOC*), currency code (Compustat field *CURCDD*), price quotation unit (Compustat field *QUNIT*), daily total return factor (Compustat field *TRFD*), and adjustment factor (issue)-cumulative by ex-date (Compustat field *AJEXDI*). These data are available in the Global Security Daily library (Compustat file *GSECD*). We compute monthly stock returns using the last day of each month with a positive closing price. Furthermore, we impose that days must have a price code status (Compustat field *PRCSTD*) equal to 3 (high, low, and close prices) or 10 (prices as reported). All currency-denominated variables are converted to U.S. dollars by using Compustat daily exchange rates (Compustat file *EXRT_DLY*).

For Canadian stock market data sourced from Compustat North America, we follow the same approach used for Compustat Global data but complement daily data from the Security Daily library (Compustat file *SECD*) with monthly data from the Security Monthly library (Compustat file *SECM*). We proceed in this manner because there are missing data for some stocks in the daily data files (Bessembinder et al. 2019). Hence, we use daily data when available, and monthly data otherwise. To be consistent in the use of adjustment factors, we impose that the time-series of prices and returns for each security relies only on either daily or monthly data. We also allow the price code status (Compustat field *PRCSTD*) to take value 4 (bid, ask, average/last volume close) in addition to 3 and 10. This takes into account the fact that Compustat North America has historically presented prices as the average of bid and ask prices.

CRSP firms and stocks are identified via *PERMCO* and *PERMNO*, respectively. Compustat firms and stocks are identified via *GVKEY* and the combination *GVKEY-IID*, respectively.

A.2 Sample Selection

We use the following Compustat files to obtain the variables based on which we apply sample selection filters: *R_COUNTRY*, *R_EX_CODES*, *SECURITY*, and *COMPANY*. The CRSP files we use for the same purpose are *CCMXPF_LNKUSED* and *MSENAMES*.

We retain a given country and stock exchange in the database if either Bessembinder et al. (2019) or Chaieb et al. (2021) do so. This ensures that minor stock exchanges with low trading volume are not included in the analysis. In addition, to ensure a minimum standard of ESG coverage in each country, we only retain countries for which there is at least one rater covering at least 10 stocks during the entire sample period. This criterion excludes Jordan and Oman from the sample.

Within each stock exchange, we only retain primary issues of common stock. Primary issues are identified using the primary issue tags (Compustat fields *PRICAN* and *PRIROW*, and CRSP field

ULINKPRIM). In case of a tie (e.g., a security is recorded as a primary security in both CRSP and Compustat over an overlapping listing period), we select the issue with the longest listing period and, in case of another tie, we select the issue traded in the headquarter country (Compustat field *LOC*).

We select common stocks as follows. For U.S. stocks, this amounts to selecting stocks with the CRSP sharecode 10, 11, and 12 (CRSP field *SHRCD*). For Compustat stocks, we retain common stock by imposing the following filters:

- the issue type (Compustat field *TPCI*) must be "0".
- the issue description (Compustat field *DSCI*) is not allowed to contain the symbol "%". These securities are likely to be be preferred stocks with fixed dividends.
- the *DSCI* keyword filters used in Chaieb et al. (2021) to remove non-common stock are applied. These filters expand on the extensive filters involving hundreds of keywords detailed in Griffin, Kelly, and Nardari (2010) and help screen out duplicates, depository receipts, preferred stock, warrants, debt, unit trusts, expired securities, and investment vehicles. In addition, we add the following keywords:
  - applied to Canadian securities (restricted, subordinated voting, non-voting): "RESTRICTD", "NVTG", "SVTG", "NON-VTG"
  - applied to Sri Lankan securities (non-voting shares): "(NON-VTG)", "NON-VTG" ,"NVTG", "(NON-VOTING)"
  - applied to Peruvian securities (investment shares similar to preferred shares): "INVT SHS"
  - applied to Australian securities (removes one specific investment fund): "AUSTRALIAN EQUITIES STRONG B"
  - applied to all securities (removes investment trust): "UNTS INVESTMENT", "UNTS TRUST"
- we exclude securities whose business description (Compustat field *BUSDESC*), company name (Compustat fields *CONM* and *CONML*), or Global Industry Classification Standard (GICS)

(Compustat fields *GSUBIND* and *GIND*) allows us to identify investment funds and trusts. We convert all text to lowercase letters, substitute ".", ",", and ";" for white spaces, delete leading and trailing white spaces at the beginning and end of sentences in *CONM*, *CONML*, add leading and trailing white spaces at the beginning and end of sentences to *BUSDESC*, and remove repeated white spaces. We also remove repeated white spaces whenever these are created at any step of the filtering process below. We proceed as follows:

1. we flag a security to be a Real Estate Investment Trusts if *GSUBIND* is "40401010" or *GIND* is "404020".
2. we identify securities to be a fund or trust from the their business description (*BUSDESC*). As a first step, we avoid false positives by removing occurrences of company names (*CONM* and *CONML*) from *BUSDESC*.
3. since in some instances "and" is written as "&", we convert all the "&" to "and" in the three fields mentioned in the previous step and repeat that step a second time.
4. we add leading and trailing spaces to every keyword listed below as well as to each entry of *CONM* and *CONML*. We create a copy of these variables and name them *CONM2* and *CONML2*. This latter step ensures that first and last words are detected. We then transform the following keywords in *CONM2* and *CONML2* into blanks and remove repeated blanks: "tel aviv", "ltd", "inc", "corp", "plc ici", "plc", "sa", "limited", "berhad", "ab", "tbk", "ag", "co", "as", "bhd", "spa", "pcl", "nv", "asa", "corporation", "pjsc", "s.a.", "se", "group", "oyj", "a/s", "a.s.", "(publ)", "cv", "holdings", "s.a", "(pt)", "ltd.", 'saog", "nl", "kk", "akcyjna", "inc.", "s.p.a.", "sirketi", "kgaa", "pt", "jsc", "s.p.a", "n.v.", "(the)", "bruxelles", "sas", "modaraba", "saa", "c.v.", "ojsc", "co.ltd.", "madrid", "lima", "a s", "s a", "oy", "london", "sca", "holding", "milano", "incorporated", "c v", "n v", "b v", "s a b", "s. a. b", "torino", "a.s", "roma", "berlin", "muenchen", "anonyme", "stockholm", "wien", "n.v", "zuerich", "hamburg", "zug", "psc", "sab", "warszawa", "augsburg", "bv", "lp", "na", "s.a.a", "sa/nv", "schaffhausen", "stuttgart", "tas", "gmbh", "llc", "incorporation", "p.s.c.", "(bbva)", "abp", "coltd", "corp.", "helsinki", "porto", "santiago", "vevey", "enterprises", "duesseldorf", "casablanca", "groupe", 'aktiebolag", "aktiengesellschaft", "bern", "bilbao", "bologna", "cagliari", "baar", "essen", "frankfurt", "cva", "hldg", "hldgs", "k.k", "k k", "ptc",

"s.a.a.", "s/a", "esp", "sarl", "(pakistan)", "marseille", "geneve", 's.a.s.", "c.v", "(bo)", "(bs)", "(gbr)", "(new)", "(re)", "-old", "lld", "ltd)", "grundbesitz-ag", "corp)", "co.ltd", "s.a.o.g.", "saog", "p.l.c", "grp", "lt", "ind".

5. We remove occurrences of company names (*CONM2* and *CONML2*) from *BUSDESC*. Note that we add a leading and trailing space to each of these words and remove all repeated white spaces before applying this filter.

6. we transform "-" in *CONM2* and *CONML2* into white spaces and remove repeated white spaces and apply the previous filter again.

7. we remove "(" and ")" from *CONM*, *CONML*, and *BUSDESC*, and flag securities as funds or trusts if either *CONML* or *CONM* contains at least one of the following keywords: " fund", " trust", "venture capital trust", " vct", or " reit".

8. we remove the following expressions from *BUSDESC*: " fund advisors"," fund managers"," fund benchmarks"," fund raisings"," fund administrations"," fund transfers"," fund services"," fund products"," fund sponsors"," fund plan sponsors"," fund corps"," fund companies"," fund groups"," trust advisors"," trust banks"," trust managers"," trust sponsors"," reit managers"," fund advisor"," fund manager"," fund benchmark"," fund raising"," fund administration"," fund transfer"," fund service"," fund product"," fund sponsor"," fund plan sponsor"," fund corp"," fund company"," fund group"," trust advisor"," trust bank"," trust manager"," trust sponsor"," reit manager"," feeder"," multi-asset"," multi asset"," balanced"," fixed income"," self-managed"," public"," publicly owned"," publicly-owned"," closed ended"," closed end"," closed-ended"," closed-end"," close ended"," close end"," close-ended"," close-end"," opened ended"," opened end"," opened-ended"," opened-end"," open ended"," open end"," open-ended"," open-end".

9. we define str1 = " is a "," was a "," as a "," is an "," was an "," as an "," specializes in "," operates as a "," operates as an " and str2 = "property trust ","property investment trust ","property fund ","property investment fund ","private equity trust ","private equity investment trust ","private equity fund ","private equity investment fund ","venture capital trust ","venture capital investment trust ","venture capital fund ","venture capital investment fund ","real estate trust ","real estate investment trust ","real estate

fund ","real estate investment fund ","interval fund ","investment trust ","etf ","reit ","vct ","unit trust ","unit investment trust ","split capital fund","split investment fund","split capital trust ","split capital investment trust ","exchange traded fund ","exchange-traded fund ","exchange-traded-fund ","exchange traded-fund ","currency fund ","fund ","mutual fund ","equity fund ","equity investment fund ","equity mutual fund ","hedge fund ","equity hedge fund ","traded fund ". We flag securities as investment funds or trusts if *BUSDESC* contains any of the pattern sequences {str1,str2}.

10. we flag securities if *BUSDESC* contains any of the following: " fund invests"," fund prefers to invest"," fund engages"," fund operates"," fund employs"," fund was formerly known"," fund replicates"," fund seeks to invest"," trust invests"," trust prefers to invest"," trust engages"," trust operates"," trust employs"," trust was formerly known"," trust replicates"," trust seeks to invest'.

11. we flag securities if *BUSDESC* contains any of the following in the first 101 characters " reit"," an investment trust"," real estate investment trust") and does not contain any of the following prior to the latter string: " by reit"," by an investment trust"," by real estate investment trust"," for reit"," for an investment trust"," for real estate investment trust"," of reit"," of an investment trust"," of real estate investment trust"," to reit"," to an investment trust"," to real estate investment trust"," through reit"," through an investment trust"," through real estate investment trust".

12. we remove all securities flagged as investment trusts or funds except for banks. Banks are identified if either (i) if *CONM* and *CONML* contains the word "bank" at the beginning or end, or (ii) if *CONM* or *CONML* contains "trust & banking", "trust and banking", or "securities co", or (iii) *GGROUP* equals "4010" and the security was flagged as an investment trust or fund in step (6).

A.3 Data Cleaning

We apply the the manual corrections listed in Chaieb et al. (2021) whenever they are applicable to our dataset. We employ the following corrections listed verbatim in Chaieb et al. (2021):

1. "The number of shares outstanding (*CSHOC*) is off by a factor 100 for the last two days of

June 2004. We then correct the number of shares." Filter applied to *GVKEY-IID* 149822-01C.

2. "The adjustment factor *AJEXDI* does not adjust for the 0.0513-to-1 stock split on May 20th, 2015. We remove the stock for this month." Filter applied to *GVKEY-IID* 208536-01W.

3. "There are errors caused by the change of currency to the Euro for these three European stocks. We remove them for January 1999." Filter applied to *GVKEY-IID* 103255-01W, 210759-01W, 240641-01W.

4. "In January 2005, there is an error in the adjustment factor ($AJEXDI$) when the currency changed. Other stocks' prices ($PRCCD$) and $AJEXDI$ adjust. This stock $PRCCD$ adjusts, but not its *AJEXDI*. We remove it for this month." Filter applied to *GVKEY-IID* 284439-01W.

5. "This Chilean stock has erratic and infrequent quotes before January 2004. There are price spikes on days with unavailable volumes, but classified as "prices as reported" ($PRCSTD$=10). There are no quotes on these days on Bloomberg. We remove infrequent returns before January 2004." Filter applied to *GVKEY-IID* 202022-01W.

6. "This Canadian stock is delisted on January 1st 2017, there is a spike in the price on December 30th, 2016, and the time series ends on December 2nd, 2016, on Bloomberg. We remove it for December 2016. CSXF is also missing the total return adjustment for the 100-to-1 conversion on November 1st, 2013, which creates a 100+% return. We remove it for November 2013." Filter applied to *GVKEY-IID* 185208-01C.

In addition, we correct decimal errors in the data sourced from Compustat. An example of a decimal error is the sequence of stock prices "9.12", "912.0", "9.08". The decimal error is that "912.0" is off by a factor of 100. As in Bessembinder et al. (2019), we repair such decimal errors that persist up to three consecutive periods.

The algorithm to correct decimal errors is described below. We first apply the algorithm to the time series of $QUNIT$ and *TRFD*. We then compute stock-split adjusted prices (in U.S. dollars) as $PRCD_{adj} = PRCCD \times FX \times QUNIT^{-1} \times AJEXDI^{-1}$. FX is the exchange rate. We also compute prices not-adjusted for stock splits using the same formula but omitting*AJEXDI*. We do the same for local prices by using the same formulas but omitting FX. We then apply the algorithm again on these time series of prices. Dividends-adjusted prices are then computed as $Price = PRCD_{adj} \times TRFD$. Market capitalization is computed as $MKTCAP = PRCD_{adj} \times CSHOC_{adj}$, where we define $CSHOC_{adj} = CSHOC \times AJEXDI$. Monthly stock returns, RET, are computed as ratios of consecutive end-of-month prices. The algorithm consists of the following steps:

- define the mapping $m(X(t); \Delta_1, \Delta_2) : X(t) \rightarrow \frac{X(t+\Delta_1)}{X(t+\Delta_2)}$. where X(t) denotes the value of a given time-series in month-year t and Δ_1 and Δ_2 are parameters taking an integer value between -3 and 3.

- define N to be the largest positive integer such that $5 \times 10^{N-1} < min(m(X(t); 0, -1), m(X(t); 0, 1))$. If $m(X(t); 0, -1) > 5 \times 10^{N-1}$ and $m(X(t); 0, 1) > 5 \times 10^{N-1}$ we substitute $X(t) \times 10^{-N}$ for $X(t)$. Intuitively, this step corrects decimal errors that last for one period only by dividing $X(t)$ by 10 if both $m(X(t); 0, -1)$ and $m(X(t); 0, 1)$ lie in the interval [5,50), by 100 if both $m(X(t); 0, -1)$ and $m(X(t); 0, 1)$ lie in the interval [50,500), and so on.

- define N to be the largest positive integer such that $\frac{1}{5 \times 10^{N-1}} > max(m(X(t); 0, -1), m(X(t); 0, 1))$. If $m(X(t); 0, -1) < \frac{1}{5 \times 10^{N-1}}$ and $m(X(t); 0, 1) < \frac{1}{5 \times 10^{N-1}}$ we substitute $X(t) \times 10^{N}$ for $X(t)$. Intuitively, this step corrects decimal errors that last for one period only by multiplying $X(t)$ by 10 if both $m(X(t); 0, -1)$ and $m(X(t); 0, 1)$ lie in the interval [1/5,1/50), by 100 if both $m(X(t); 0, -1)$ and $m(X(t); 0, 1)$ lie in the interval [1/50,1/500), and so on.

- define N_1 to be the largest positive integer such that $5 \times 10^{N_1-1} < min(m(X(t); 0, -1), m(X(t); 0, 2))$. If $m(X(t); 0, -1) > 5 \times 10^{N_1-1}$, $m(X(t); 0, 2) > 5 \times 10^{N_1-1}$, and $|m(X(t); 0, 1) - 1| < 30\%$, then we flag $X(t)$ as a potential first observation with decimal errors in a sequence of two observations with decimal errors.

- define N_2 to be the largest positive integer such that $5 \times 10^{N_2-1} < min(m(X(t+1); 0, -2), m(X(t+1); 0, 1))$. If $m(X(t+1); 0, 1) > 5 \times 10^{N_2-1}$ and $m(X(t+1); 0, -2) > 5 \times 10^{N_2-1}$ then we flag $X(t+1)$ as a potential second observation with decimal errors in a sequence of two observations with decimal errors.

- if both observations $X(t)$ and $X(t+1)$ are flagged, we divide them by 10^{N_1} and 10^{N_2}, respectively.

- define N_1 to be the largest positive integer such that $\frac{1}{5\times 10^{N_1-1}} > max(m(X(t);0,-1), m(X(t);0,2))$. If $m(X(t);0,-1) < \frac{1}{5\times 10^{N_1-1}}$, $m(X(t);0,2) < \frac{1}{5\times 10^{N_1-1}}$, and $|m(X(t);0,1)-1| < 30\%$, then we flag $X(t)$ as a potential first observation with decimal errors in a sequence of two observations with decimal errors.

- define N_2 to be the largest positive integer such that $\frac{1}{5\times 10^{N_1-1}} > max(m(X(t+1);0,1), m(X(t+1);0,-2))$. If $m(X(t+1);0,1) < \frac{1}{5\times 10^{N_2-1}}$ and $m(X(t+1);0,-2) < \frac{1}{5\times 10^{N_2-1}}$ then we flag $X(t+1)$ as a potential second observation with decimal errors in a sequence of two observations with decimal errors.

- if both observations $X(t)$ and $X(t+1)$ are flagged, we multiply them by 10^{N_1} and 10^{N_2}, respectively.

- define N_1 to be the largest positive integer such that $5\times 10^{N_1-1} < min(m(X(t);0,-1), m(X(t);0,3))$. If $m(X(t);0,-1) > 5\times 10^{N_1-1}$, $m(X(t);0,3) > 5\times 10^{N_1-1}$, $|m(X(t);0,1)-1| < 30\%$, and $|m(X(t);0,2)-1| < 30\%$, then we flag $X(t)$ as a potential first observation with decimal errors in a sequence of two observations with decimal errors.

- define N_2 to be the largest positive integer such that $5\times 10^{N_2-1} < min(m(X(t+1);0,-2), m(X(t+1);0,2))$. If $m(X(t+1);0,2) > 5\times 10^{N_2-1}$, $m(X(t+1);0,-2) > 5\times 10^{N_2-1}$, and $|m(X(t+1);0,1)-1| < 30\%$ then we flag $X(t+1)$ as a potential second observation with decimal errors in a sequence of two observations with decimal errors.

- define N_3 to be the largest positive integer such that $5\times 10^{N_3-1} < min(m(X(t+2);0,-3), m(X(t+2);0,1))$. If $m(X(t+2);0,1) > 5\times 10^{N_3-1}$ and $m(X(t+2);0,-3) > 5\times 10^{N_3-1}$, then we flag $X(t+2)$ as a potential third observation with decimal errors in a sequence of two observations with decimal errors.

- if observations $X(t)$, $X(t+1)$, and $X(t+2)$ are flagged, we divide them by 10^{N_1}, 10^{N_2}, and 10^{N_3}, respectively.

- define N_1 to be the largest positive integer such that $\frac{1}{5\times 10^{N_1-1}} > max(m(X(t);0,-1), m(X(t);0,3))$. If $m(X(t);0,-1) < \frac{1}{5\times 10^{N_1-1}}$, $m(X(t);0,3) < \frac{1}{5\times 10^{N_1-1}}$, $|m(X(t);0,1)-1| < 30\%$, and

$|m(X(t);0,2)-1| < 30\%$, then we flag $X(t)$ as a potential first observation with decimal errors in a sequence of two observations with decimal errors.

- define N_2 to be the largest positive integer such that $\frac{1}{5\times10^{N_2-1}} > max(m(X(t+1);0,-2), m(X(t+1);0,2))$. If $m(X(t+1);0,2) < \frac{1}{5\times10^{N_2-1}}$, $m(X(t+1);0,-2) < \frac{1}{5\times10^{N_2-1}}$, and $|m(X(t+1);0,1)-1| < 30\%$ then we flag $X(t+1)$ as a potential second observation with decimal errors in a sequence of two observations with decimal errors.

- define N_3 to be the largest positive integer such that $\frac{1}{5\times10^{N_3-1}} > max(m(X(t+2);0,-3), m(X(t+2);0,1))$. If $m(X(t+2);0,1) < \frac{1}{5\times10^{N_3-1}}$ and $m(X(t+2);0,-3) < \frac{1}{5\times10^{N_3-1}}$, then we flag $X(t+2)$ as a potential third observation with decimal errors in a sequence of two observations with decimal errors.

- if observations $X(t)$, $X(t+1)$, and $X(t+2)$ are flagged, we multiply them by 10^{N_1}, 10^{N_2}, and 10^{N_3}, respectively.

After correcting decimal errors, we apply a series of filters to remove remaining errors in Compustat which are relatively more frequent in small and illiquid stocks, stocks with low share prices, and during the first months after a stock starts being covered in the database. We proceed as follows:

- we compute the average number of daily observations with positive trading volume for each stock-month. We average this number across months for each stock. We exclude stocks in the lowest 3% of the distribution of the latter metric.

- we drop stocks for which the adjustment factor $AJEXDI$ ever takes value of 0.

- we drop stocks if they experience (i) changes in quotation units ($QUNIT$), (ii) without contemporaneous changes in currency code ($CURCCD$), and (iii) contemporaneous changes in $PRCD_{adj}$ larger than 50% in absolute value.

- we delete the remaining time-series for any stock if its non-adjusted (for stock splits and dividends) share price drops below $U.S. 0.01.

- we delete the remaining time-series for any stock if its market capitalization drops below $U.S. 1 million.

- if the return data contains gaps for more than 11 months, we set the first month after the data resumes to missing.

- for all stocks other than those listed in China, we define an observation as a jump if one of the following conditions is met:

  1. $\frac{CSHOC_{adj}(t)}{CSHOC_{adj}(t-1)} \geq 5$ and $\frac{MKTCAP(t)}{MKTCAP(t-1)} \geq 2.5$
  2. $\frac{CSHOC_{adj}(t)}{CSHOC_{adj}(t-1)} \leq 0.2$ and $\frac{MKTCAP(t)}{MKTCAP(t-1)} \leq 0.4$

- for Chinese stocks, we define an observation as a jump if one of the following conditions is met:

  1. $\frac{CSHOC_{adj}(t)}{CSHOC_{adj}(t-1)} \geq 50$ and $\frac{MKTCAP(t)}{MKTCAP(t-1)} \geq 25$
  2. $\frac{CSHOC_{adj}(t)}{CSHOC_{adj}(t-1)} \leq 0.2$ and $\frac{MKTCAP(t)}{MKTCAP(t-1)} \leq 0.4$

- we delete the observation in which the jump occurs and all the observations thereafter if the jump occurs during the lesser of the first 24 months of data for the stock or during the first 20% of the observations for that stock (whichever is smaller). For jumps which do not match these criteria, we proceed as follows:

  1. if the jump at time t is reversed by another jump at time s: we replace $CSHOC_{adj}$ by $min(CSHOC_{adj}(t-1), CSHOC_{adj}(s))$ for observations in interval $[t, s-1]$.
  2. if the jump is not reversed, we adjust $CSHOC_{adj}$ to be smaller for observations before or after the jump. For upward jumps occurring at time t, we multiply $CSHOC_{adj}(t)$ and all observations thereafter by $\frac{CSHOC_{adj}(t-1)}{CSHOC_{adj}(t)}$. For downward jumps occurring at time t, we multiply all observations up to and including $CSHOC_{adj}(t)$ by $\frac{CSHOC_{adj}(t)}{CSHOC_{adj}(t-1)}$.

- we identify cases in which market capitalization jumps in a manner inconsistent with the behavior of the time-series of returns and shares outstanding. We define an observation as a jump if one of the following conditions holds:

  1. $\frac{MKTCAP(t)}{MKTCAP(t-1)-1} > 9$ and $RET < 2$
  2. $\frac{MKTCAP(t)}{MKTCAP(t-1)-1} < -0.9$ and $RET > -0.5$

- we delete the observation in which the jump occurs and all the observations thereafter if the jump occurs during the lesser of the first 24 months of data for the stock or during the first 20% of the observations for that stock (whichever is smaller). For jumps which do not match these criteria, we proceed as follows:

  1. if the jump at time t is reversed by another jump at time s: we adjust market capitalization by replacing $MKTCAP$ by $min(MKTCAP(t-1) \times (1 + RETX(t)), \frac{MKTCAP(s)}{(1+RETX(s))})$ for observations in interval $[t, s-1]$. $RETX$ is computed as RET using the time-series of prices $PRCD_{adj}$.
  2. if the jump is not reversed, we adjust $MKTCAP$ to be smaller for observations before or after the jump. For upward jumps occurring at time t, $MKTCAP(t)$ and all observations thereafter are multiplied by $\frac{MKTCAP(t-1)}{MKTCAP(t)} \times (1 + RETX(t))$. For downward jumps occurring at time t, all observations up to and including $MKTCAP(t)$ are multiplied by $\frac{MKTCAP(t)}{MKTCAP(t-1) \times (1+RETX(t))}$.

- we delete observations for which the changes in returns are inconsistent with the changes in market capitalization. These observations are those that satisfy one of these conditions:

  1. $\frac{MKTCAP(t)}{MKTCAP(t-1)} - 1 < 0.5$ and $RET(t) > 0.8$
  2. $\frac{MKTCAP(t)}{MKTCAP(t-1)} - 1 > -0.5$ and $RET(t) < -0.8$

- we delete stock-months that are amongst the first three months for that stock provided they satisfy one of the following conditions:

  1. $\frac{PRCD_{adj}(t)}{PRCD_{adj}(t-1)} > 10$ and $\frac{MKTCAP(t)}{MKTCAP(t-1)} > 10$
  2. $\frac{PRCD_{adj}(t)}{PRCD_{adj}(t-1)} < 10^{-1}$ and $\frac{MKTCAP(t)}{MKTCAP(t-1)} < 10^{-1}$

- we exclude stock-months with fewer than five daily observations of positive closing prices in that month or the month before. We do not apply this filter to delisting months. The identification of delisting months is explained below.

- we delete stocks for which there are fewer than six months of stock return data available.

A.4 Industry Filters

We exclude observations in both CRSP and Compustat with NAICS (North American Industry Classification System) industry code of 525 corresponding to "Funds, Trusts, and Other Financial Vehicles".

A.5 Delisting returns

The delisting returns for CRSP stocks are adjusted following Beaver, McNichols, and Price (2007). This adjustment is similar to the adjustment of Shumway (1997), who suggests assuming a delisting return of -30% when the delisting return is missing in the database and the delisting reason is poor performance. The reason for this choice is the mean delisting return estimated in their sample is -30%. It is also common to assume a delisting return of -100% when the delisting return is missing and the delisting reason is not performance related (Bali et al. 2016). Beaver et al. (2007), however, find that average delisting return varies substantially depending on the delisting reason. Hence, instead of assuming a single delisting return, we follow Beaver et al. (2007) and estimate mean replacement values by type of delisting.

Following Bessembinder et al. (2019), delisting return for Compustat stocks is set to -30% if one of the following conditions holds:

1. the delisting reason is bankruptcy or liquidation (Compustat field $DLRSN = 02$ or $DLRSN = 03$).
2. the stock-month has active security status (Compustat field $SECSTAT$) but the stock does not experience any changes in prices during at least 12 months before the end of the sample. We consider the stock to be inactive during these months in which the price does not change and delete them. The last remaining month for these stocks is considered a delisting month. Note that, in this last step, we use the price in local currency because the price in foreign currency changes mechanically due to exchange rate fluctuations.

Figure IA.1. Correlations of E subratings across rating agencies

This figure summarizes the full sample correlations of E subratings across rating agencies for each geographic region and for the world. The first column reports the average correlations across all rating-pairs. The table also shows the minimum and maximum correlation across rating-pairs. Since the RepRisk rating is negatively correlated with the other ratings, we also report averages and minima of correlations across all rating-pairs other than RepRisk. The correlations are based on all firms in our sample and on the full sample period from January 2001 until December 2020.

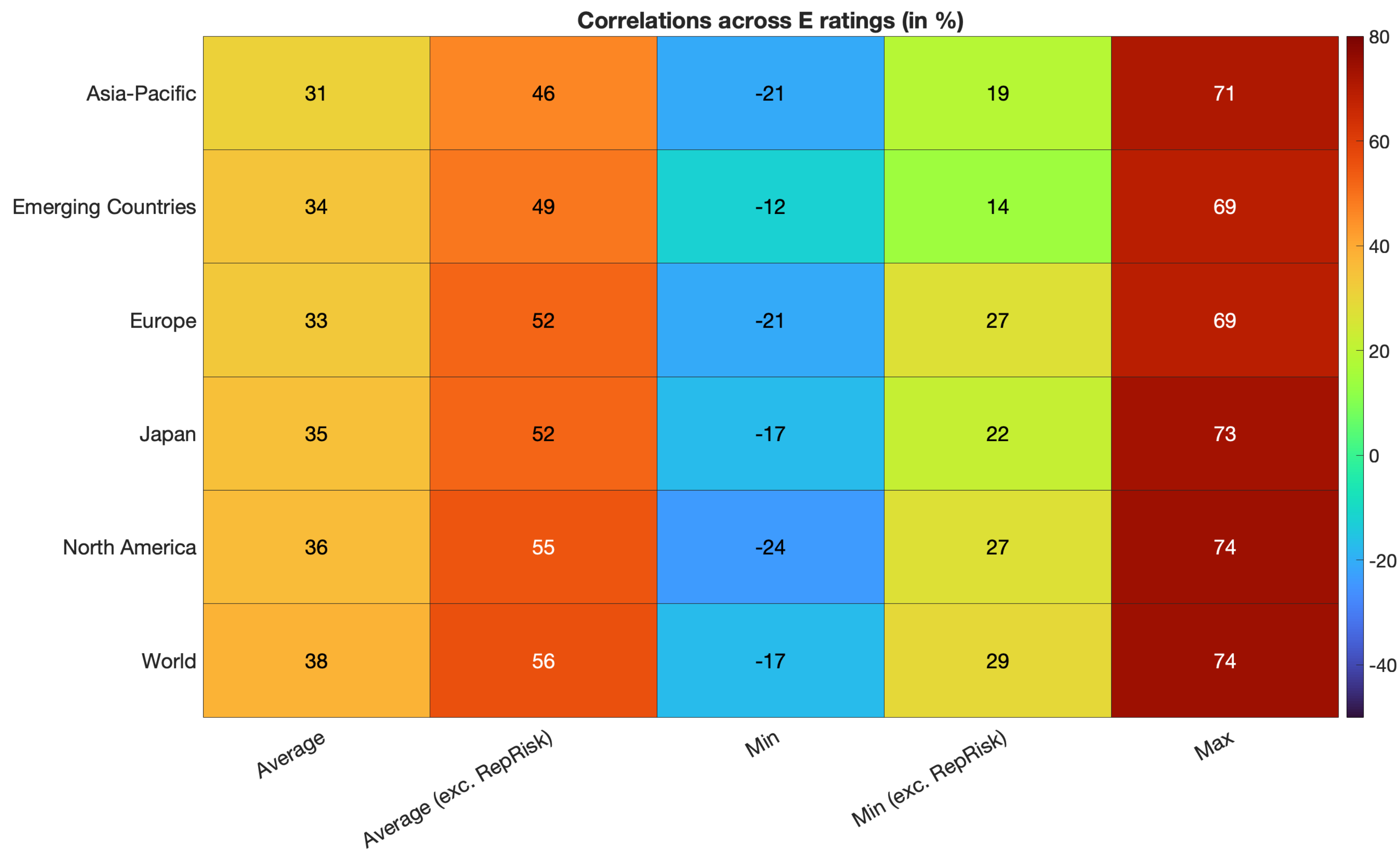


Figure IA.2. Correlations of S subratings across rating agencies

This figure summarizes the full sample correlations of S subratings across rating agencies for each geographic region and for the world. The first column reports the average correlations across all rating-pairs. The table also shows the minimum and maximum correlation across rating-pairs. Since the RepRisk rating is negatively correlated with the other ratings, we also report averages and minima of correlations across all rating-pairs other than RepRisk. The correlations are based on all firms in our sample and on the full sample period from January 2001 until December 2020.

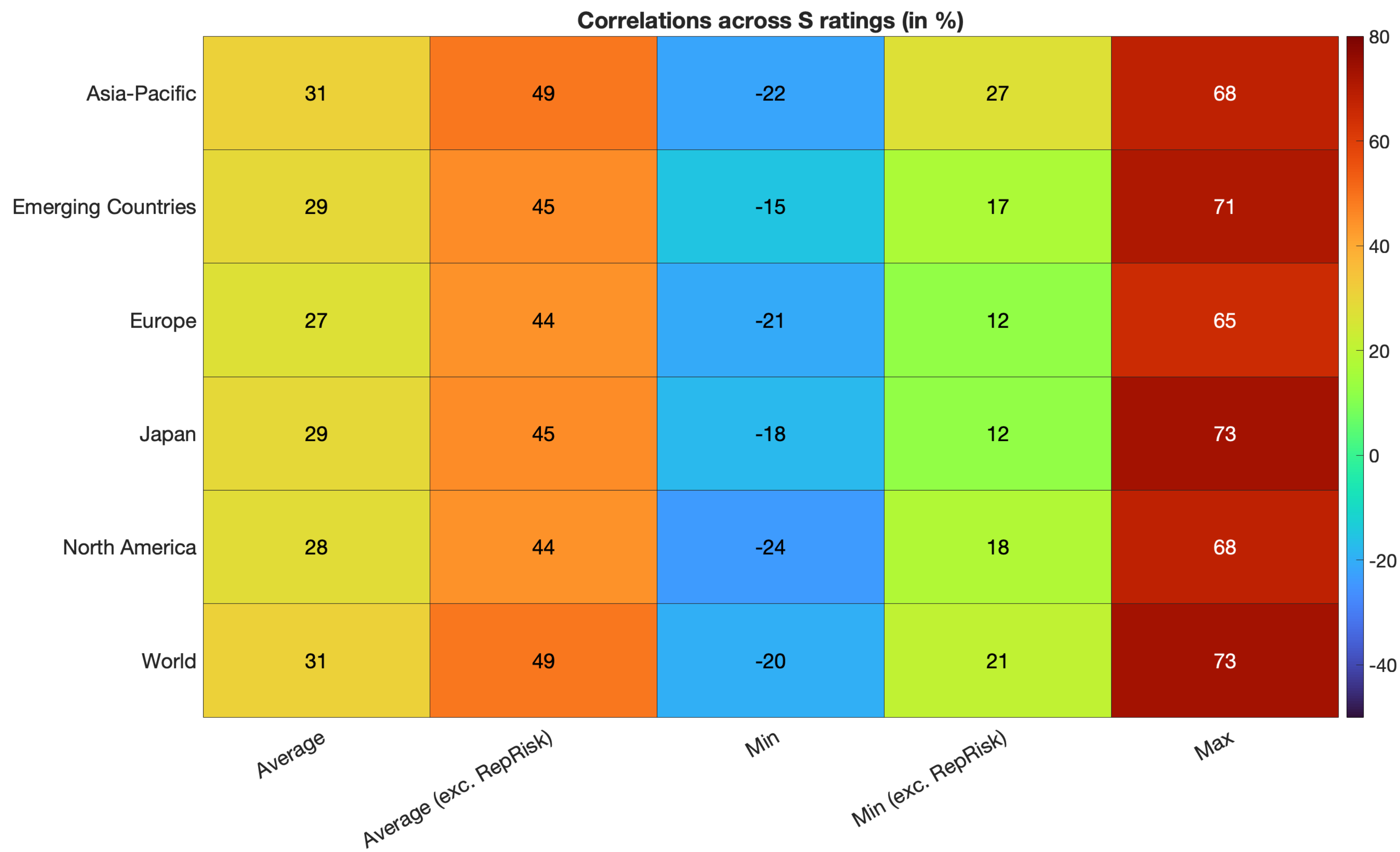


Figure IA.3. Correlations of G subratings across rating agencies

This figure summarizes the full sample correlations of G subratings across rating agencies for each geographic region and for the world. The first column reports the average correlations across all rating-pairs. The table also shows the minimum and maximum correlation across rating-pairs. Since the RepRisk rating is negatively correlated with the other ratings, we also report averages and minima of correlations across all rating-pairs other than RepRisk. The correlations are based on all firms in our sample and on the full sample period from January 2001 until December 2020.

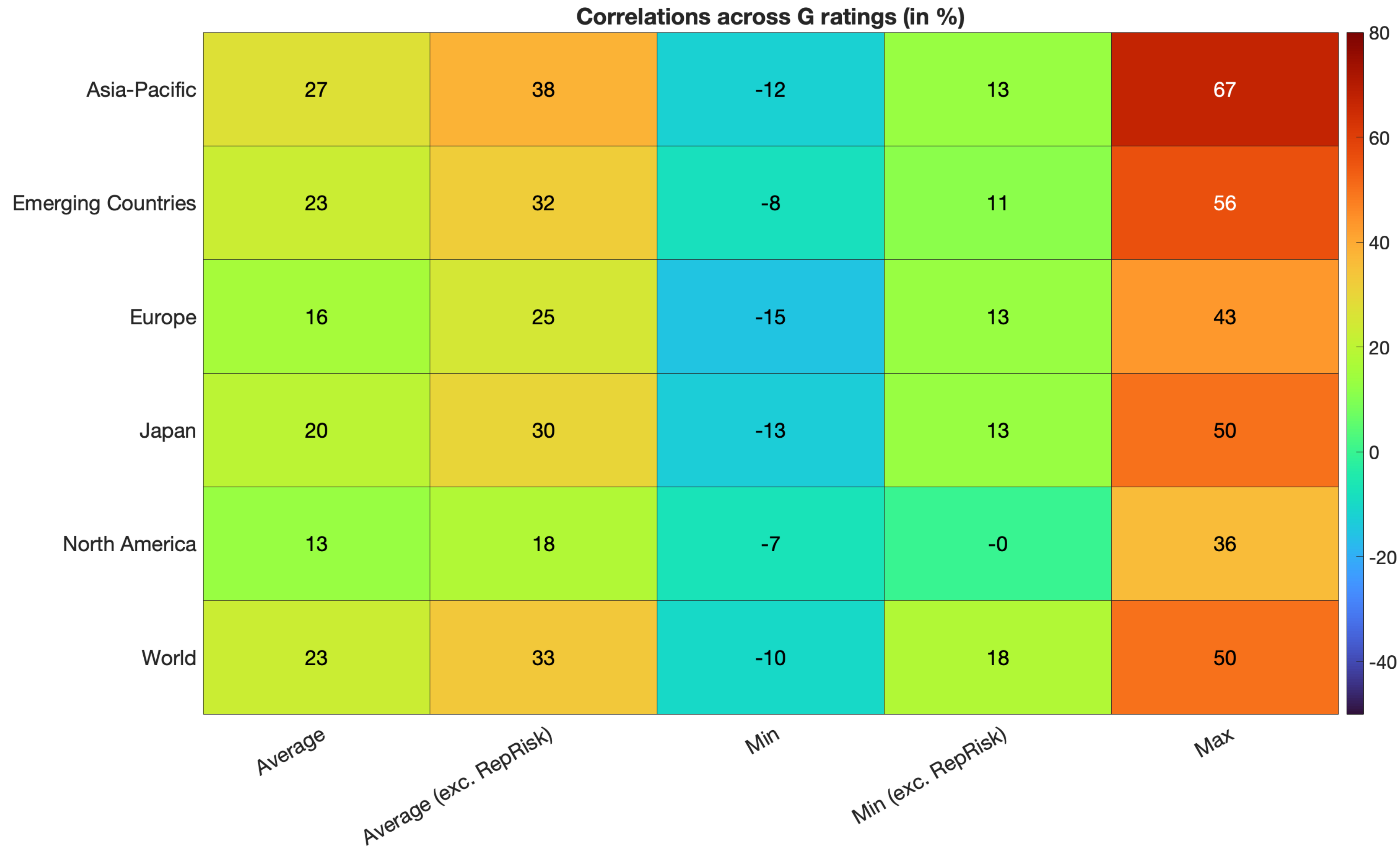
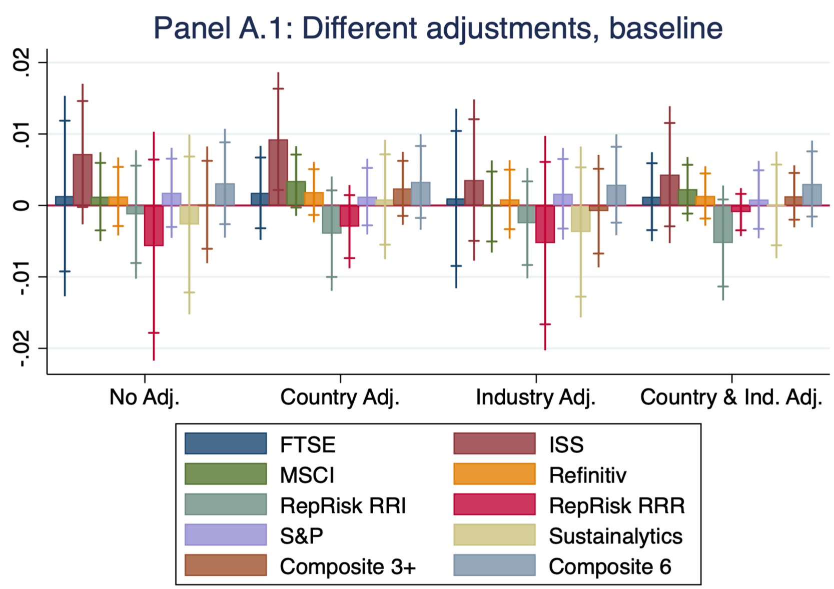


Figure IA.4. Value-weighted Fama-Macbeth regressions

This figure shows the results of several variations of the baseline Fama-Macbeth regressions in Table 1 of monthly stock returns on lagged ESG ratings using the following ten different ratings: FTSE, ISS, MSCI IVA, Refinitiv, RepRisk (RRI), RepRisk (RRR), S&P Global, Sustainalytics, Composite 3+, and Composite 6. The panels on the right display the results from value-weighted Fama-Macbeth regressions. The panels on the left display our baseline results presented in Figure 3 for comparison purposes. In comparison to Figure 3, we adjust the y-axis of the left panels as needed to align with the scale of the right panels. The weights used in the value-weighted regressions are based on market capitalization at the end of the previous month. Each panel presents, in bars, the coefficient on the different ESG ratings obtained in separate regressions that include the full set of control variables from Table 1. Confidence intervals are depicted as whiskers around the point estimates. The longer (shorter) whiskers represent 99% (95%) confidence intervals. Confidence intervals use Newey and West (1987) standard errors with automatic lag selection. Panel A shows the results when using country- and/or industry-adjusted ESG ratings, where the baseline results based on unadjusted ratings are included at the left-hand side of the panel. Panel B shows the results for the E, S, and G subratings. Panel C shows the results for different geographic regions. Panel D shows the results for the the period after the Paris Agreement (January 2016 to December 2020), using unadjusted, country- and/or industry-adjusted ESG ratings. Panel E shows the results for the E subrating separately for the U.S. and the rest of the world excluding U.S. (RoW) controlling either for the full set of controls or for the firm characteristics that correspond to the factors of the Fama and French (2015) five-factor model plus momentum (FF6). Panel F reports the results for ESG momentum, defined as the most recent change in ESG ratings during the previous 12 months, using unadjusted, country- and/or industry-adjusted ESG ratings. Panels G and H show the results for ESG downgrades and upgrades, respectively, under different adjustments and using the global sample. Panels I (Panel J) reports the results for worst-in-class and best-in-class ESG strategies using the global sample (U.S. subsample). The sample period is from January 2001 to December 2020, but the starting dates vary across raters and regions.



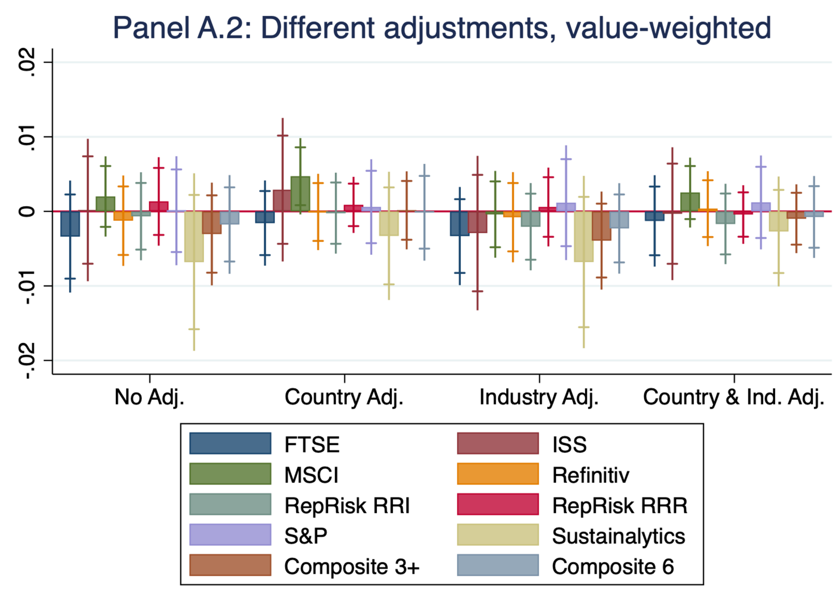


Figure IA.4 - continued

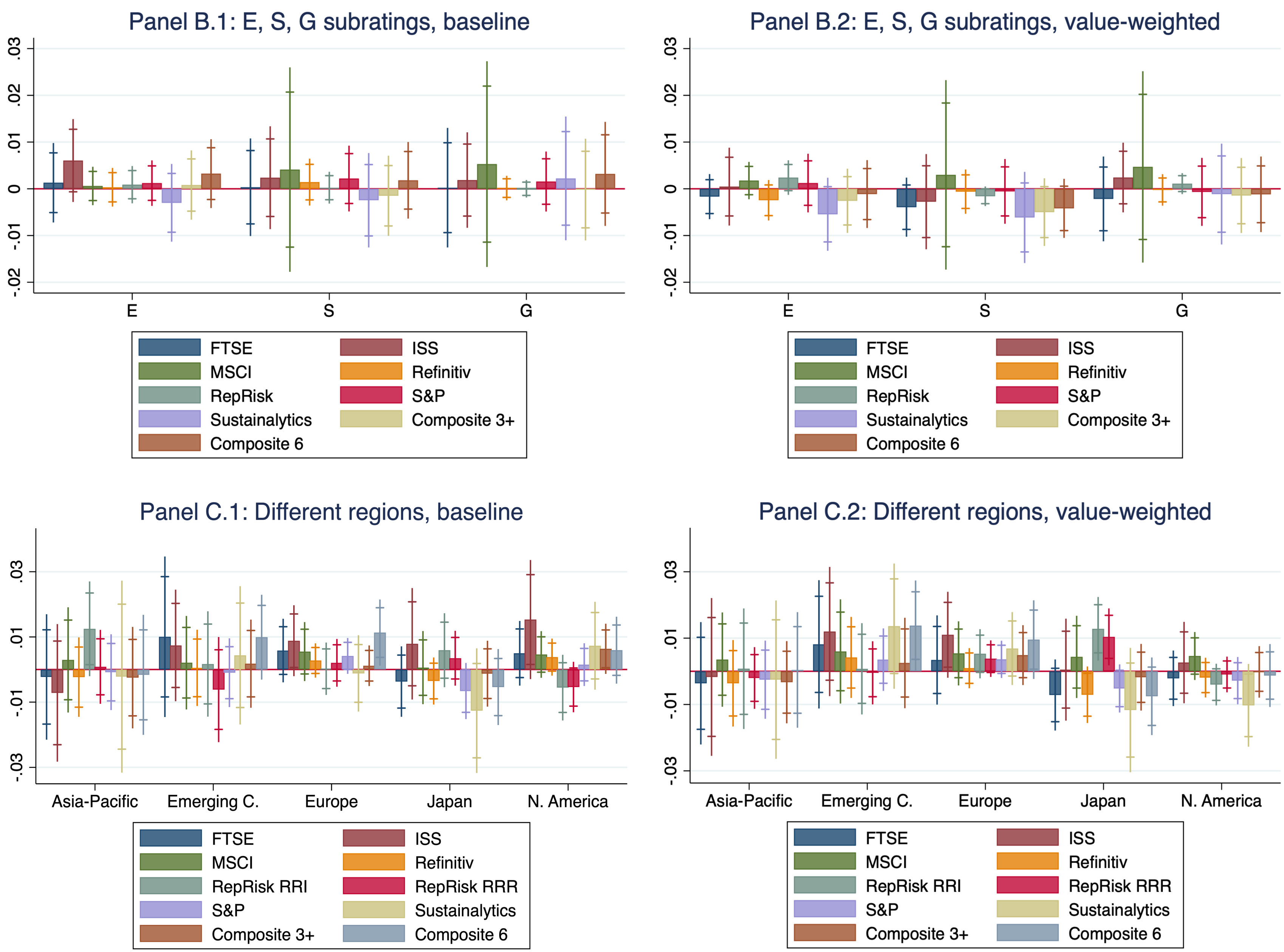
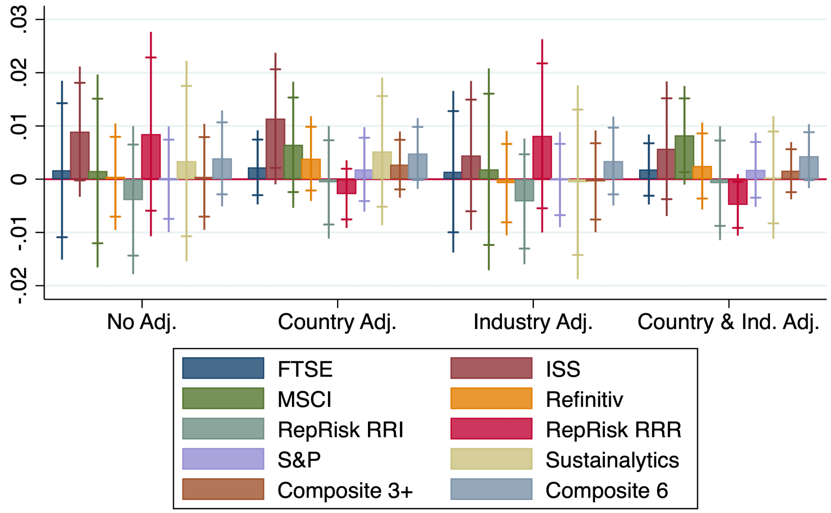
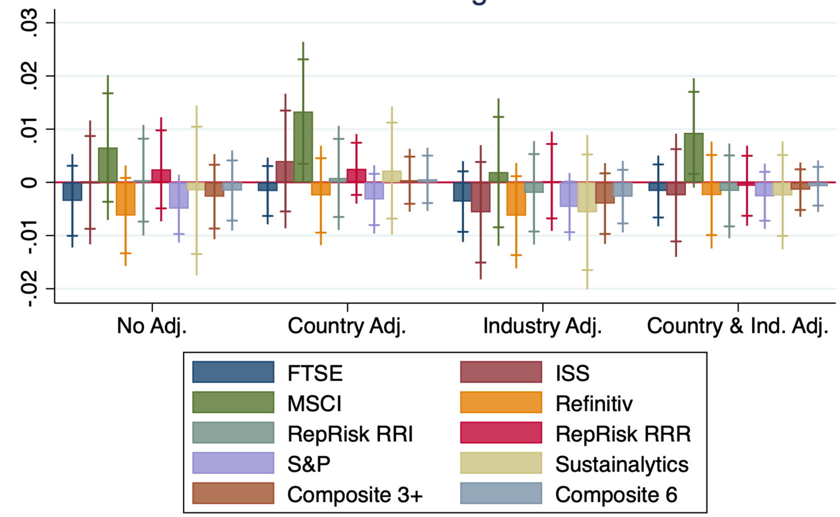


Figure IA.4 - continued

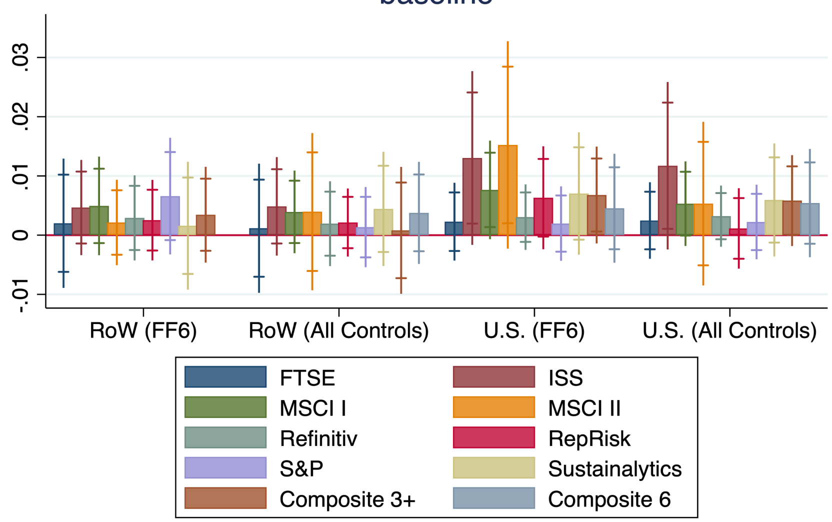
Panel D.1: After Paris Agreement, different adjustments, baseline



Panel D.2: After Paris Agreement, different adjustments, value-weighted



Panel E.1: E subrating, U.S. vs. Rest of World (RoW), baseline



Panel E.2: E subrating, U.S. vs. Rest of World (RoW), value-weighted

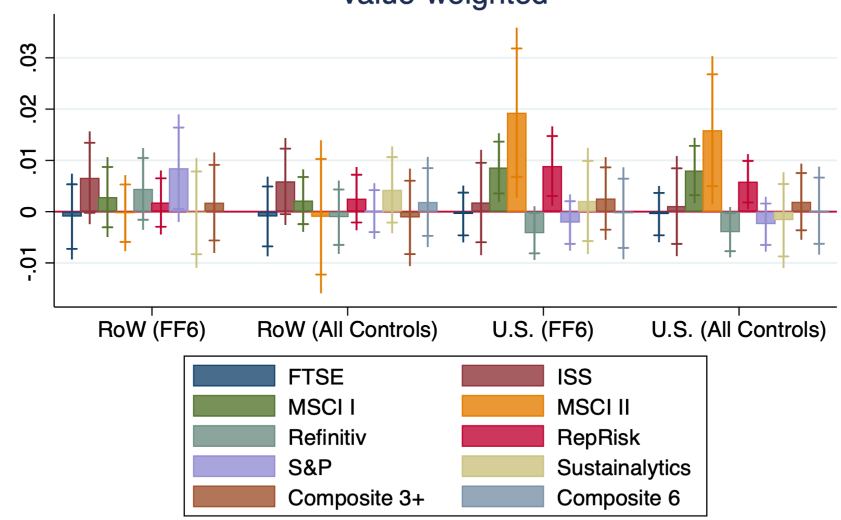
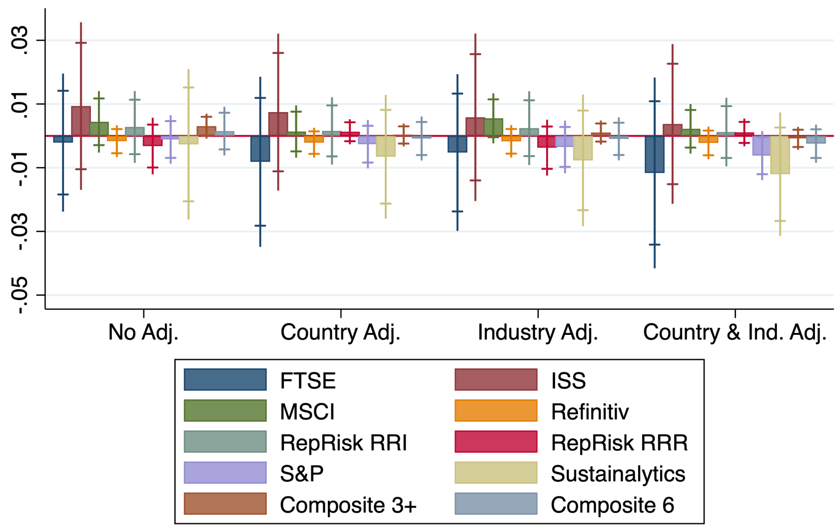
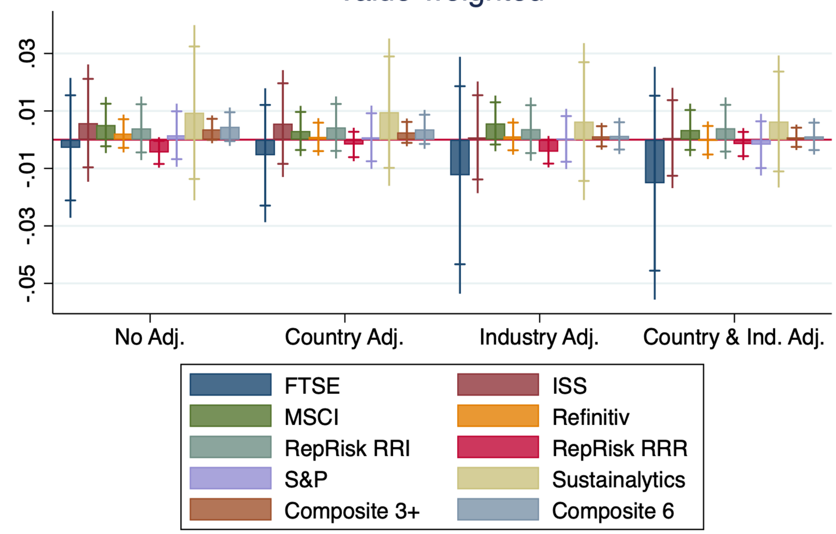


Figure IA.4 - continued

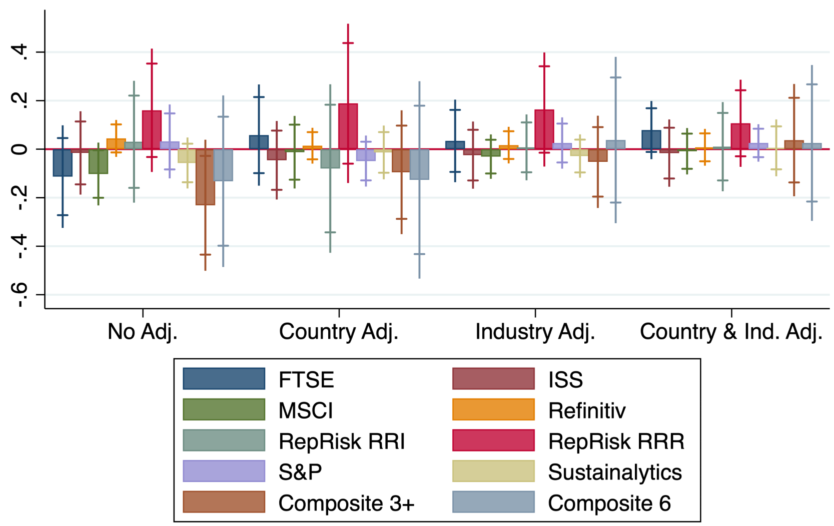
Panel F.1: ESG momentum, different adjustments, baseline



Panel F.2: ESG momentum, different adjustments, value-weighted



Panel G.1: ESG downgrades, different adjustments, baseline



Panel G.2: ESG downgrades, different adjustments, value-weighted

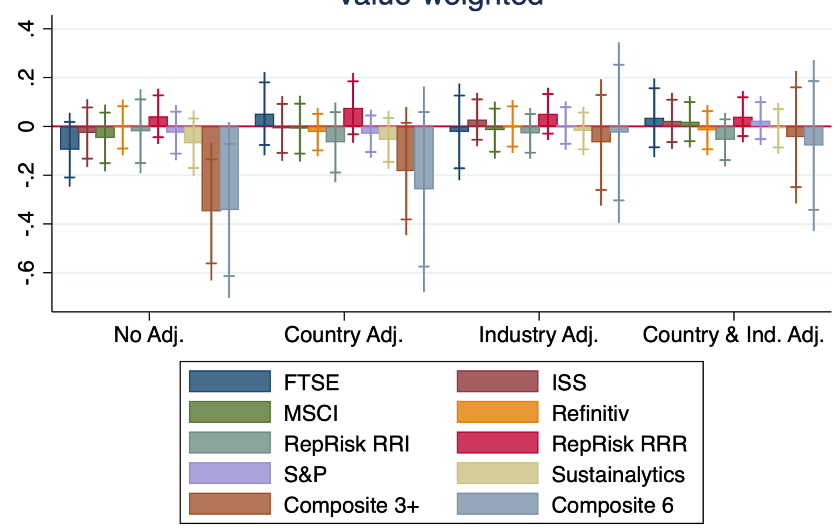


Figure IA.4 - continued

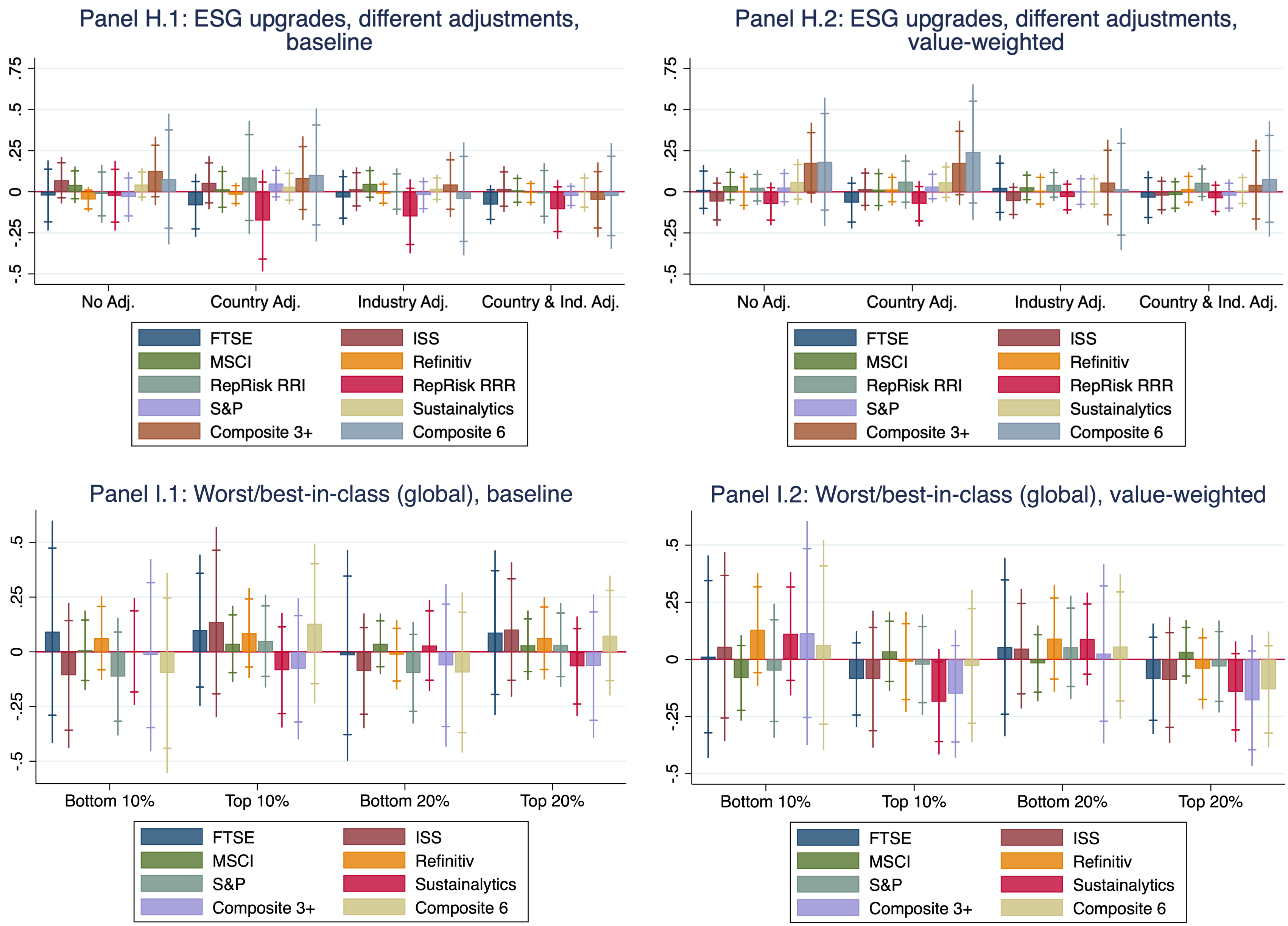
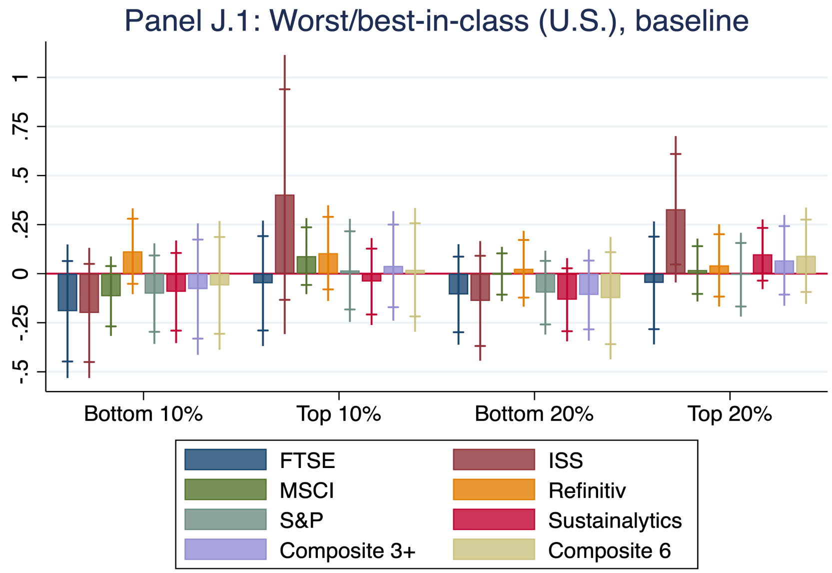


Figure IA.4 - continued



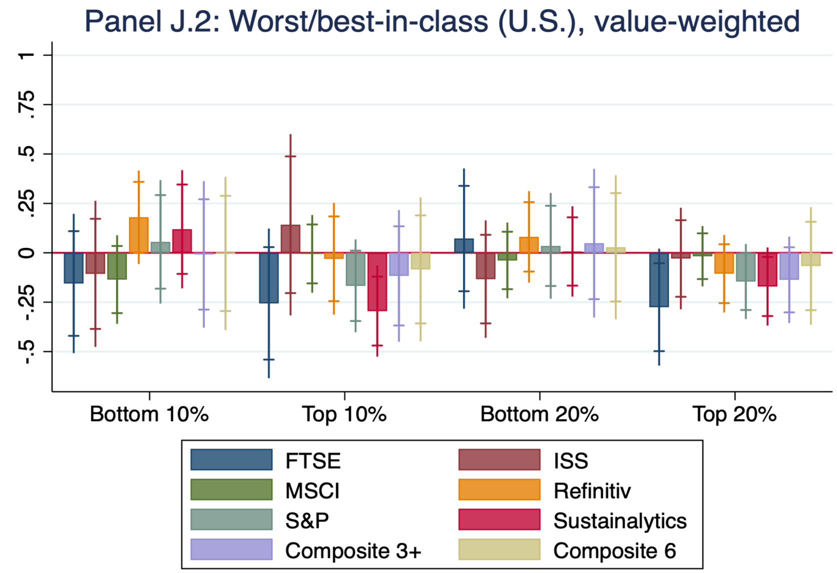


Table IA.1. Variable definitions and data sources

This table provides the definitions and data sources of the variables used throughout the paper.

ESG Variables	
ESG ratings	ESG ratings and E, S and G subratings are sourced from FTSE, ISS, MSCI IVA, Refinitiv, RepRisk, S&P Global, and Sustainalytics. In the case of RepRisk, we use the two different ratings provided by the rater, RRR (RepRisk rating) and RRI (RepRisk index). The RRR adjusts the RRI by taking into account firms' risk exposures in the countries and sectors where ESG incidents take place. Since the RRR is available in letter format, we convert it to a numerical format on a scale between 0 and 100 via linear interpolation, such that AAA corresponds to 100 and D corresponds to 0. We transform the RRI ratings provided by RepRisk by converting them to negative values and adding 100, so that higher ratings capture lower reputational risk exposure to ESG issues. In the case of ISS, the rater provides a joint S and G rating ($SOGGOV$) in lieu of a S subrating before March 2017. We treat $SOCGOV$ scores as S subratings before that date. Some raters provide overall ESG ratings with and without industry adjustments. In these cases we use the ratings without industry adjustments, as recommended by Pástor et al. (2021), to account for the possibility that ESG investor preferences are priced in on the basis of unadjusted ratings. However, we also create industry- and/or country-level adjusted ratings for every rater by, at each point in time and each stock, subtracting from each rating the average rating of country and/or industry peers in that month. When adjusting ratings at the country-level, we require that the rater covers at least 10 stocks in that country during the entire sample period to compute the adjusted rating. Since Sustainalytics introduced a new ESG dataset with a fundamentally different methodology in 2018, we opt for the legacy dataset ending in 2018 because this is the dataset that was de facto used by investors during the overwhelming majority of the sample period. Since Sustainalytics ratings change at most once a year and are highly autocorrelated, we use the last ratings available in the legacy dataset as of October 2018 to predict subsequent stock returns from November 2018 onwards.
Composite ratings	We construct three composite ESG ratings by averaging ratings across raters: *Composite 2+*, *Composite 3+*, and *Composite 6*. We exclude RepRisk in the construction of composite raters because RepRisk is negatively correlated with, and conceptually different from, the other six raters. We convert ratings to percentile ranks each month before averaging. *Composite 2+* covers stocks that are rated by at least two of the six raters. *Composite 3+* covers stocks that are rated by at least three of the six raters. *Composite 6* covers stocks that are rated by all six raters.

Table IA.1 - continued

ESG variables	
ESG momentum	Most recent change in ESG ratings that occurred in the preceding 12 months. In the case of Refinitiv and S&P Global, which report ratings at yearly frequency, ESG momentum is the year-on-year change in ratings. ESG momentum is missing during the first 12 months after a stock enters the dataset.
ESG upgrades and downgrades	We define ESG upgrades (downgrades) as positive (negative) ESG momentum. For composite raters (*Composite 3+*, and *Composite 6*), we define upgrades (downgrades) as the proportion of raters that upgrade (downgrade) a stock.
Low ESG uncertainty	Dummy variable equal to one if a stock is among the 20% of stocks with lowest ESG uncertainty in a given month. ESG uncertainty is the standard deviation of ESG ratings across all six raters other than RepRisk. Ratings are converted to percentile ranks in each month before computing the standard deviation. ESG uncertainty is missing if fewer than two raters rate a stock in a given month.
Stock returns and characteristics	
Returns	For U.S. stocks, we use the monthly stock returns in the CRSP field *RET*. For non-U.S. stocks we compute monthly stock returns in U.S. dollars using Compustat daily data and the formula $\frac{PRCCD_t \times FX_t \times QUNIT_{t-1} \times AJEXDI_{t-1} \times TRFD_t}{PRCCD_{t-1} \times FX_{t-1} \times QUNIT_t \times AJEXDI_t \times TRFD_{t-1}}$. $PRCCD_t$ denotes the closing stock price at time t. $TRFD_t$, $AJEXDI_t$ and $QUNIT_t$ adjust for dividends, stock-splits, and quotation units. FX_t is the exchange rate from local currency to U.S. dollars. We use the last day of each month with a non-zero closing price to compute monthly returns. Returns are adjusted for delisting as described in the delisting returns subsection of the Internet Appendix.
Beta	Regression slope of a regression of monthly excess stock returns on country-specific market excess returns over the previous 24 months. We impose that at least 12 months of valid excess return observations must be available during the 24-month period. We use one-month Treasury bill rates to compute excess returns. U.S. market excess returns and Treasury bill rates are retrieved from Kenneth French's data library. For the remaining countries, we compute market returns as the weighted average of stock market returns in each country and month. The weights are proportional to the market capitalization of each stock.
Size	Natural logarithm of market capitalization in millions of U.S. dollars. For U.S. stocks, following Bali et al. (2016), market capitalization is the absolute value of the product of end-of-month stock price (CRSP field *ALTPRC*) and the number of shares outstanding (CRSP field *SHROUT*). For non-U.S. stocks, market capitalization is computed analogously using the data fields described in the definition of the variable *Returns*. Following Fama and French (1992), we compute market capitalization as of June of each year and hold that value constant until May of the following year.

Table IA.1 - continued

Stock returns and characteristics	
B/M	Natural logarithm of the ratio of book value of equity to market capitalization. The book value of equity is computed as the sum of the Compustat items book value of stockholder's equity (*SEQ*) and deferred taxes and investment tax credit (*TXDITC*), minus the book value of preferred stock which is defined as either the redemption value (*PRTKRV*), liquidating value (*PSTKL*), or par value (*PSTK*) as available (in this order). If none of these measures is available we assume the book value of preferred stock is zero. If *SEQ* is missing, we replace it by the sum of book value of equity (*CEQ*) and book value of preferred stock if available. If unavailable, we use the difference between total assets (*AT*) and total liabilities (*LT*). If *TXDITC* is missing, we replace it by the sum of deferred taxes (*TXBD*) and investment tax credit (*ITCB*), where *ITCB* is set to zero if missing. If *TXDITC* is still missing, we set it to zero. Market capitalization is computed as described in the definition of the variable *Size*. Following Fama and French (1992), the B/M ratio in June of year *t+1* through May of year *t+2* is computed using market capitalization measured as of the end of calendar year *t* and book value of equity measured at the end of the fiscal year ending in calendar year *t*.
Investment	Percentage change in total assets (Compustat item *AT*).
Gross profitability	Revenues (Compustat item *REVT*) minus costs of goods sold (Compustat item *COGS*) divided by assets (Compustat item *AT*).
Momentum	Momentum in month *t* is defined as the cumulative return in U.S. dollars over the 11-month period between months *t-2* and *t-12*. We require at least nine months of available return data during the 11-month period.
Leverage	Ratio of total debt (Compustat items *DLTT*+*DLC*) to total assets (Compustat item *AT*). Following Jensen, Kelly, and Pedersen (2023), we allow either *DLTT* or *DLC* to be missing but not both. Negative values of *DLTT* and *DLC* are set to missing if negative, following Iliev and Welch (2010).
Volatility	Annualized standard deviation of monthly stock returns in U.S. dollars over the previous 12 months.
Tangibility	Property, plant, and equipment (Compustat item *PPENT*) divided by total assets (Compustat item *AT*).
R&D	R&D spending (Compustat item *XRD*) divided by total assets (Compustat item *AT*). When R&D is missing, it is assumed zero.

Table IA.1 - continued

Environmental and social norms	
Environmental norms	Index measuring environmental norms in a country. The index ranges between zero and one. Higher values mean stronger environmental norms. It is based on survey responses to questions related to: (i) unpaid work related to environment, conservation, and animal rights; (ii) active/inactive membership in environmental organizations; (iii) whether it is important to a person to look after the environment; (iv) whether a person would forgo part of their income for the environment; and (v) whether protecting the environment has priority over economic growth. Responses across questions are combined following the methodology of Welzel (2013). The data is sourced from the Integrated Values Survey and uses data from World Values Survey (Waves 4-7) and European Values Survey (Waves 4 and 5). The index is updated as new survey data for each country becomes available (i.e., not every year). Since data from different waves exhibit very high autocorrelation (Dyck et al. 2019), we compute the average value of the index over the sample period for each country to maximize coverage and cancel out noise.
Social norms	Index measuring social norms in a country. The index ranges between zero and one. Higher values mean stronger social norms. It is based on survey responses to questions related to: (i) autonomy: whether independence and imagination are important child qualities; (ii) gender equality: a) men should have more right to jobs than women; b) men make better political leaders than women do; c) university is more important for a boy than for a girl; d) men make better business executives than women do; (iii) voice: assign first, second, or no priority to the goals of (a) protecting freedom of speech; (b) giving people more say in important government decisions; (c) giving people more say about how things are done at their jobs and in their communities; (iv) freedom: how acceptable respondents find (a) divorce; (b) abortion; and (c) homosexuality. Responses across questions are combined following the methodology of Welzel (2013). The data is sourced from the Integrated Values Survey and uses data from World Values Survey (Waves 4-7) and European Values Survey (Waves 4 and 5). The index is updated as new survey data for each country becomes available (i.e., not every year). Since data from different waves exhibit very high autocorrelation (Dyck et al. 2019), we compute the average value of the index over the sample period for each country to maximize coverage and cancel out noise.

Table IA.1 - continued

Environmental and social norms	
Associational activity	Index measuring the extent to which individuals in a country actively participate in recreational, humanitarian, and environmental associations. The index ranges between zero and one. Higher values mean more associational activity. The index is based on World Values Survey data and aggregated at the country-level. The index is invented and constructed by Welzel (2013) and data is sourced from the Welzel replication file (Welzel 2015) available in the online GESIS Data Archive administered by the Leibniz-Institut für Sozialwissenschaften. For further details refer to Welzel (2015).
Social movement activity	Index measuring the extent to which peaceful social movement activities (petitions, demonstrations, and boycotts) are part of a country's culture. The index ranges between zero and one. Higher values mean more social movement activity. The index is based on World Values Survey data and aggregated at the country-level. The index is invented and constructed by Welzel (2013) and data is sourced from the Welzel replication file (Welzel 2015) available in the online GESIS Data Archive administered by the Leibniz-Institut für Sozialwissenschaften. For further details refer to Welzel (2015).
Schmidt Political Orientation index	Index that measures the political orientation of a country's government based on the percentage of cabinet positions held by different parties, taking into account the number of days in office in a given year. The index ranges from one to five. A value of one means hegemony of right-wing and centre parties (left-wing parties account for 0% of cabinet positions). A value of two means dominance of right-wing and centre parties (left-wing parties account for 33.33% or less of cabinet positions). A value of three means balance of power between left and right wing parties (left-wing parties account for between 33.33% and 66.67% of cabinet positions). A value of four means dominance of social-democratic and other left-wing parties (left-wing parties account for 66.67% or more of cabinet positions). A value of five means hegemony of social-democratic and other left-wing parties (left-wing parties account for 100% of cabinet positions). Sourced from the Comparative Political Data Set (Armingeon, Engler, Leemann, and Weisstanner 2023).
Votes green parties	Share of votes obtained by parties classified as green in the most recent national parliament election in a given country. Sourced from the Comparative Political Data Set (Armingeon et al. 2023).
Votes left parties I	Share of votes obtained by parties classified as left socialist, green, or feminist in the most recent national parliament election in a given country. Sourced from the Comparative Political Data Set (Armingeon et al. 2023).

Table IA.1 - continued

Environmental and social norms	
Votes left parties II	Share of votes obtained by parties classified as left socialist, green, feminist, communist, or post-communist in the most recent national parliament election in a given country. Sourced from the Comparative Political Data Set (Armingeon et al. 2023).
Votes left parties III	Share of votes obtained by parties classified as social democrat, left socialist, green, feminist, communist, or post-communist in the most recent national parliament election in a given country. Sourced from the Comparative Political Data Set (Armingeon et al. 2023).
Votes non-right parties I	Share of votes obtained by parties not classified as conservative or right-wing in the most recent national parliament election in a given country. Sourced from the Comparative Political Data Set (Armingeon et al. 2023).
Votes non-right parties II	Share of votes obtained by parties not classified as conservative, right-wing, or religious in the most recent national parliament election in a given country. Sourced from the Comparative Political Data Set (Armingeon et al. 2023).
Votes non-right parties III	Share of votes obtained by parties not classified as liberal, conservative, right-wing, or religious in the most recent national parliament election in a given country. Sourced from the Comparative Political Data Set (Armingeon et al. 2023).
Left-wing government I	Share of cabinet posts of social democratic and other left-wing parties as a percentage of total cabinet posts in a given country-year. The computation takes into account the number of days in office in a given year. Sourced from the Comparative Political Data Set (Armingeon et al. 2023).
Left-wing government II	Share of seats in parliament of social democratic and other left-wing parties in government as a percentage of total parliamentary seats held by all government parties in a given country-year. The computation takes into account the number of days in office in a given year. This is intended as a measure of relative power of social democratic and other left-wing parties within the government. Sourced from the Comparative Political Data Set (Armingeon et al. 2023).
Left-wing government III	Share of seats in parliament of social democratic and other left-wing parties in government as a percentage of total parliamentary seats. The computation takes into account the number of days in office in a given year. Sourced from the Comparative Political Data Set (Armingeon et al. 2023).

Table IA.1 - continued

ESG disclosure standards	
Mandatory ESG disclosure	Indicator that equals one starting from January of the year after a country introduced mandatory ESG disclosure, and zero otherwise. If ESG disclosure is not introduced all at once, we require that mandatory E, S, and G disclosure is present for the indicator to be one. Sourced from Krueger et al. (2024).
Mandatory E disclosure	Indicator that equals one starting from January of the year after a country introduced mandatory environmental (E) disclosure, and zero otherwise. Sourced from Krueger et al. (2024).
Mandatory S disclosure	Indicator that equals one starting from January of the year after a country introduced mandatory social (S) disclosure, and zero otherwise. Sourced from Krueger et al. (2024).
Mandatory G disclosure	Indicator that equals one starting from January of the year after a country introduced mandatory governance (G) disclosure, and zero otherwise. Sourced from Krueger et al. (2024).
Mandatory ESG disclosure (all at once)	Indicator that equals one starting from January of the year after a country introduced all-at-once implementation of ESG disclosure, and zero otherwise. The all-at-once implementation means that mandatory disclosure on E, S, and G was introduced at the same time. Sourced from Krueger et al. (2024).
Mandatory ESG disclosure (government mandated)	Indicator that equals one starting from January of the year after a government institution in a country introduced mandatory ESG disclosure, and zero otherwise. A government institution can be a ministry, the parliament, a securities regulator, or a similar institution. If ESG disclosure is not introduced all at once, we require that mandatory E, S, and G disclosure is present for the indicator to be one. Sourced from Krueger et al. (2024).
Mandatory ESG disclosure (full compliance)	Indicator that equals one starting from January of the year after a country introduced mandatory ESG disclosure on a full-compliance basis (not on a comply-or-explain basis), and zero otherwise. If ESG disclosure is not introduced all at once, we require that mandatory E, S, and G disclosure is present for the indicator to be one. Sourced from Krueger et al. (2024).

Table IA.1 - continued

Strictness of environmental regulations	
Environmental Performance index	Index that measures environmental performance in terms of the extent to which countries achieve environmental policy targets. The index ranges between zero and 100. Higher values mean better environmental performance. It is based on a set of 40 performance indicators covering climate change performance, environmental health, and ecosystem validity. Higher index values indicate better environmental performance in a country. The index is released every two years and we update index values accordingly. Sourced from the Yale Center for Environmental Law.
Environmental Democracy index	Index that measures the extent to which countries pass legally binding laws and practices that improve transparency, accountability, and citizen engagement in environmental decision-making. The index ranges between zero and three. Higher values mean more transparency and accountability on environmental issues. It is based on three pillars: (i) right to access information on environmental quality and problems; (ii) right to participate meaningfully in decision-making; and (iii) right to demand enforcement of environmental laws and compensation for harm. Sourced from the World Resources Institute (WRI).
Environmental Policy Stringency index	Index that measures the extent to which countries' environmental policies put an implicit or explicit price on environmental harmful behavior. The index ranges between zero and six. Higher values mean more environmental stringency. It is based on 13 environmental policy instruments, mostly related to climate and air pollution. This index is administered by the Organisation for Economic Co-operation and Development (OECD).
Employment Laws index	Index that captures the rigidity of labor regulations in a country. The index ranges between zero and one. Higher values mean more rigid labor regulation. It is based on three dimensions: (i) restrictions placed on alternative employment contracts to capture the extent to which these contracts are used to bypass regular labor provisions; (ii) conditions of employment (e.g., mandatory payment for non-working days, minimum wage legislation, flexibility of working time requirements); and (iii) job security (e.g., dismissal procedures and severance payment). This index is sourced from Botero et al. (2004).

Table IA.1 - continued

Strictness of social regulations	
Labor Regulation index	Index that measures the rigidity of labor regulation in a given country. The index ranges between zero and ten. Higher values mean more flexible labor regulation. It is based on six dimensions: (i) hiring regulations and minimum wage; (ii) hiring and firing regulations; (iii) centralized collective bargaining; (iv) hours regulations; (v) mandated cost of worker dismissal; (vi) conscription. Sourced from the Fraser Institute.
Employment Protection Legislation I	Index of labor market flexibility. The index ranges between five and ten. Higher values mean more flexible labor regulation. This is based on three components: (i) individual dismissal of workers with regular contracts (EPR) which accounts for factors such as the easiness of dismissal and severance pay; (ii) additional costs for collective dismissals relative to the costs of individual dismissals (EPC); and, (iii) regulation of temporary contracts (EPT), which covers considerations such as compensation and working conditions of temporary workers. Following Edmans et al. (2024b), we define the Employment Protection Index I as the equally-weighted average of the three components. We assign 2019 index values to 2020 because the data end in 2019. Sourced from the Organisation for Economic Co-operation and Development (OECD). We use the most recent version (version 4) of the dataset.
Employment Protection Legislation II	Index of labor market flexibility. The index ranges between five and ten. Higher values mean more flexible labor regulation. This is based on three components (EPR, EPC, and EPT) defined above. Following Edmans et al. (2024b), we define Employment Protection Legislation II as the weighted average of the three components using weights $\frac{10}{21}$, $\frac{4}{21}$, and $\frac{7}{21}$ for EPR, EPC, and EPT, respectively. We assign 2019 index values to 2020 because the data end in 2019. Sourced from the Organisation for Economic Co-operation and Development (OECD). We use the most recent version (version 4) of the dataset.

Table IA.2. Summary statistics

This table reports the mean and standard deviation of stock returns and various stock characteristics (Panel A), ESG ratings (Panel B), and ESG momentum (Panel C). In the case of RepRisk, we report the RepRisk Index (RRI) and RepRisk Rating (RRR) instead of ESG ratings. Variable definitions are available in Internet Appendix Table IA.1. The table also reports the starting date, the number of unique stocks, and the number of stock-months covered by each rater and jointly across all raters (full sample).

	Full Sample		FTSE		ISS		MSCI IVA		Refinitiv	
No. Stocks	16,368		4,766		5,814		9,872		6,444	
No. Stock-Months	1,528,507		188,509		261,490		668,880		507,195	
Start Date	2001-Jan		2015-Jan		2013-Apr		2001-Jan		2003-Jul	
Variable	Mean	S.D	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Panel A: Stock Returns & Characteristics										
Returns (in % p.m.)	1.034	14.274	0.975	10.481	1.069	12.148	0.952	11.781	1.055	12.211
Beta	1.106	0.668	1.055	0.538	1.084	0.587	1.113	0.665	1.108	0.631
Size	6.844	1.914	8.711	1.372	8.348	1.723	8.103	1.496	8.239	1.559
B/M	-0.631	1.046	-0.803	1.066	-0.839	1.071	-0.814	0.976	-0.774	1.023
Investment	0.156	0.449	0.098	0.293	0.113	0.348	0.138	0.399	0.124	0.368
Gross Profitability	0.271	0.237	0.274	0.205	0.278	0.219	0.292	0.237	0.278	0.230
Momentum (in %)	10.769	51.321	4.660	31.490	5.937	36.855	8.638	39.429	8.823	40.609
Volatility (in % p.a.)	39.750	23.263	29.992	14.603	32.661	18.796	33.452	18.521	34.080	19.860
Leverage	0.224	0.179	0.235	0.167	0.242	0.170	0.227	0.172	0.228	0.168
Tangibility	0.301	0.236	0.284	0.228	0.277	0.239	0.279	0.240	0.285	0.238
R&D	0.019	0.052	0.017	0.036	0.022	0.051	0.023	0.055	0.022	0.055
Panel B: ESG Ratings										
ESG			49.533	20.569	24.894	14.635	47.428	12.690	42.098	20.675
E			41.961	27.732	22.870	17.924	48.426	20.852	34.373	28.902
S			43.306	24.364	23.617	13.264	46.524	16.821	41.613	23.713
G			64.757	20.533	39.495	17.972	53.509	18.892	49.217	22.649
Panel C: ESG Momentum										
ESG			2.257	7.179	0.776	3.243	0.212	5.724	2.225	7.307
E			1.800	11.018	0.700	4.178	0.437	9.516	2.524	10.021
S			2.661	9.600	0.655	4.152	0.277	9.547	2.459	9.030
G			2.568	10.080	0.897	6.495	0.654	12.026	1.379	13.534

Table IA.2 - continued

	RepRisk		S&P Global		Sustainalytics		Composite 3+		Composite 6	
No. Stocks	1,208,160		395,688		380,497		282,124		116,404	
No. Stock-Months	10,629		6,985		4,556		6,188		2,237	
Start Date	2007-Feb		2007-Jul		2009-Sep		2014-Jan		2015-Jan	
Variable	Mean	S.D	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Panel A: Stock Returns & Characteristics										
Returns (in % p.m.)	0.995	14.580	1.044	10.704	0.876	10.747	0.953	11.548	0.956	9.650
Beta	1.100	0.652	1.072	0.573	1.091	0.585	1.095	0.589	1.036	0.513
Size	6.731	1.973	8.655	1.375	8.514	1.351	8.492	1.340	9.302	1.107
B/M	-0.586	1.066	-0.793	1.006	-0.746	1.030	-0.838	1.051	-0.893	1.076
Investment	0.154	0.443	0.109	0.310	0.103	0.323	0.111	0.329	0.092	0.275
Gross Profitability	0.262	0.228	0.287	0.216	0.277	0.210	0.275	0.213	0.270	0.198
Momentum (in %)	10.708	52.996	8.403	36.511	8.431	35.833	5.558	34.979	4.930	28.753
Volatility (in % p.a.)	40.648	23.659	31.301	16.928	31.525	16.724	31.946	17.505	28.099	13.146
Leverage	0.230	0.179	0.238	0.164	0.238	0.167	0.239	0.169	0.245	0.160
R&D	0.315	0.235	0.293	0.232	0.296	0.241	0.284	0.238	0.283	0.231
Tangibility	0.015	0.043	0.018	0.038	0.018	0.040	0.019	0.045	0.017	0.036
Panel B: ESG Ratings										
ESG			35.784	21.609	57.660	9.906	52.152	23.184	59.270	23.112
E	92.570	20.216	31.116	26.823	55.411	13.795	51.773	22.533	59.978	21.895
S	86.766	28.000	32.903	22.314	56.443	11.509	51.323	21.722	57.520	21.881
G	87.009	30.128	41.836	20.780	63.190	10.233	52.230	19.780	55.122	19.731
RRI	93.344	11.022								
RRR	72.372	20.676								
Panel C: ESG Momentum										
ESG			0.324	8.494	0.378	2.406	49.905	14.089	50.639	12.951
E	-0.252	16.298	1.411	10.712	0.630	3.945	49.996	14.140	50.891	13.123
S	-0.996	22.899	0.166	9.547	0.443	3.936	49.806	14.069	50.499	12.950
G	-1.285	24.605	-0.361	9.842	0.280	3.675	50.085	13.492	50.472	12.096
RRI	0.007	3.907								
RRR	-1.182	8.637								

Table IA.3. Equally-weighted portfolio sorts adjusted for multiple hypothesis testing

This table summarizes the alphas of equally-weighted portfolio strategies based on ESG ratings. We sort stocks at the end of each month into high and low ESG rating portfolios and compute the return spread (high minus low) in excess of the one-month U.S. Treasury bill rate. The sorts are conducted separately for all combinations of geographic regions, ESG (sub)ratings, country and/or industry rating adjustments, and three alternative portfolio breakpoints (top and bottom 10%, 20%, or 30%). The table summarizes raw return spreads as well as alphas relative to five different popular factor models: the CAPM, the Fama and French (1993) three-factor model (*FF3*), the Carhart (1997) four-factor model (*Carhart*), the Fama and French (2015) five-factor model (*FF5*), and the five-factor model plus momentum (*FF5+MOM*)). For each combination of (sub)rating, risk model, geographic region, and rating adjustment, the table reports two numbers: x/y. Here, x denotes the number of tests where the time-series alpha is statistically significant at the 5% level or better, and y represents the number of those significant alphas that are positive. For example, a hypothetical entry "6/5" indicates that, out of all tests for that specific combination, six alphas are statistically significant, and five of those are positive. The total number of tests per cell/combination is indicated at the bottom of the table, either 27 or 24, depending on whether ESG ratings or subratings are used. Specifically, 27 (24) tests are performed when using nine (eight) ESG ratings (subratings) and three alternative breakpoints. We use the following eight ESG subrating sources to form portfolios: FTSE, ISS, MSCI IVA, Refinitiv, S&P Global, Sustainalytics, Composite 3+, and Composite 6. Additionally, the RRR RepRisk rating is used for portfolios formed on ESG ratings. We exclude the RRI/E/S/G RepRisk ratings, as they exhibit limited cross-sectional variation across firms for sorting purposes. Significance tests are corrected for multiple hypothesis testing using the procedure of Benjamini et al. (2006), applied per column (i.e., separately for each (sub)rating and each rating adjustment). The sample period is from January 2001 to December 2020, but the exact starting date for each region is chosen so that there are at least 120 stocks in that region as of the starting date. Variable descriptions are provided in the Internet Appendix.

	No Adjustment				Industry-adjusted				Country-adjusted				Country & ind. adjusted			
	ESG	E	S	G	ESG	E	S	G	ESG	E	S	G	ESG	E	S	G
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Asia-Pacific																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Emerging Countries																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0

Table IA.3 - continued

	No Adjustment				Industry-adjusted				Country-adjusted				Country & ind. adjusted			
	ESG (1)	E (2)	S (3)	G (4)	ESG (5)	E (6)	S (7)	G (8)	ESG (9)	E (10)	S (11)	G (12)	ESG (13)	E (14)	S (15)	G (16)
Europe																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF3	4/4	0/0	2/2	0/0	1/1	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	1/1	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	3/3	0/0	2/2	0/0	1/1	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Japan																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
North America																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
No. raters	9	8	8	8	9	8	8	8	9	8	8	8	9	8	8	8
No. hypotheses per cell	27	24	24	24	27	24	24	24	27	24	24	24	27	24	24	24
No. hypotheses per column	810	720	720	720	810	720	720	720	810	720	720	720	810	720	720	720

Table IA.4. Value-weighted portfolio sorts adjusted for multiple hypothesis testing, 2012/Nov-2020/Dec

This table summarizes the alphas of value-weighted portfolio strategies based on ESG ratings. We sort stocks at the end of each month into high and low ESG rating portfolios and compute the return spread (high minus low) in excess of the one-month U.S. Treasury bill rate. Stocks are value-weighted based on their market capitalization at the end of the previous month. The sorts are conducted separately for all combinations of geographic regions, ESG (sub)ratings, country and/or industry rating adjustments, and three alternative portfolio breakpoints (top and bottom 10%, 20%, or 30%). The table summarizes raw return spreads as well as alphas relative to five different popular factor models: the CAPM, the Fama and French (1993) three-factor model (*FF3*), the Carhart (1997) four-factor model (*Carhart*), the Fama and French (2015) five-factor model (*FF5*), and the five-factor model plus momentum (*FF5+MOM*)). For each combination of (sub)rating, risk model, geographic region, and rating adjustment, the table reports two numbers: x/y. Here, x denotes the number of tests where the time-series alpha is statistically significant at the 5% level or better, and y represents the number of those significant alphas that are positive. For example, a hypothetical entry "6/5" indicates that, out of all tests for that specific combination, six alphas are statistically significant, and five of those are positive. The total number of tests per cell/combination is indicated at the bottom of the table, either 27 or 24, depending on whether ESG ratings or subratings are used. Specifically, 27 (24) tests are performed when using nine (eight) ESG ratings (subratings) and three alternative breakpoints. We use the following eight ESG subrating sources to form portfolios: FTSE, ISS, MSCI IVA, Refinitiv, S&P Global, Sustainalytics, Composite 3+, and Composite 6. Additionally, the RRR RepRisk rating is used for portfolios formed on ESG ratings. We exclude the RRI/E/S/G RepRisk ratings, as they exhibit limited cross-sectional variation across firms for sorting purposes. Significance tests are corrected for multiple hypothesis testing using the procedure of Benjamini et al. (2006), applied per column (i.e., separately for each (sub)rating and each rating adjustment). The sample period is from November 2012 to December 2020, but the exact starting date for each region is chosen so that there are at least 120 stocks in that region as of the starting date. Variable descriptions are provided in the Internet Appendix.

	No Adjustment				Industry-adjusted				Country-adjusted				Country & ind. adjusted			
	ESG	E	S	G	ESG	E	S	G	ESG	E	S	G	ESG	E	S	G
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Asia-Pacific																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0	0/0	0/0	0/0
FF3	0/0	3/0	1/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	2/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	3/0	4/0	3/0	2/0	0/0	0/0	1/0	0/0	0/0	1/0	3/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	1/0	2/0	3/0	1/0	0/0	0/0	0/0	0/0	0/0	1/0	3/0	0/0	0/0	0/0	0/0	0/0
Emerging Countries																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	3/3	0/0	1/1	0/0	2/2	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	3/3	0/0	1/1	0/0	1/1	0/0	0/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/1	0/0	1/1	0/0	1/1	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/1	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0

Table IA.4 - continued

	No Adjustment				Industry-adjusted				Country-adjusted				Country & ind. adjusted			
	ESG (1)	E (2)	S (3)	G (4)	ESG (5)	E (6)	S (7)	G (8)	ESG (9)	E (10)	S (11)	G (12)	ESG (13)	E (14)	S (15)	G (16)
Europe																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF3	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Japan																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
North America																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
No. raters	9	8	8	8	9	8	8	8	9	8	8	8	9	8	8	8
No. hypotheses per cell	27	24	24	24	27	24	24	24	27	24	24	24	27	24	24	24
No. hypotheses per column	810	720	720	720	810	720	720	720	810	720	720	720	810	720	720	720

Table IA.5. Equally-weighted portfolio sorts adjusted for multiple hypothesis testing, 2012/Nov-2020/Dec

This table summarizes the alphas of equally-weighted portfolio strategies based on ESG ratings. We sort stocks at the end of each month into high and low ESG rating portfolios and compute the return spread (high minus low) in excess of the one-month U.S. Treasury bill rate. The sorts are conducted separately for all combinations of geographic regions, ESG (sub)ratings, country and/or industry rating adjustments, and three alternative portfolio breakpoints (top and bottom 10%, 20%, or 30%). The table summarizes raw return spreads as well as alphas relative to five different popular factor models: the CAPM, the Fama and French (1993) three-factor model (*FF3*), the Carhart (1997) four-factor model (*Carhart*), the Fama and French (2015) five-factor model (*FF5*), and the five-factor model plus momentum (*FF5+MOM*)). For each combination of (sub)rating, risk model, geographic region, and rating adjustment, the table reports two numbers: x/y. Here, x denotes the number of tests where the time-series alpha is statistically significant at the 5% level or better, and y represents the number of those significant alphas that are positive. For example, a hypothetical entry "6/5" indicates that, out of all tests for that specific combination, six alphas are statistically significant, and five of those are positive. The total number of tests per cell/combination is indicated at the bottom of the table, either 27 or 24, depending on whether ESG ratings or subratings are used. Specifically, 27 (24) tests are performed when using nine (eight) ESG ratings (subratings) and three alternative breakpoints. We use the following eight ESG subrating sources to form portfolios: FTSE, ISS, MSCI IVA, Refinitiv, S&P Global, Sustainalytics, Composite 3+, and Composite 6. Additionally, the RRR RepRisk rating is used for portfolios formed on ESG ratings. We exclude the RRI/E/S/G RepRisk ratings, as they exhibit limited cross-sectional variation across firms for sorting purposes. Significance tests are corrected for multiple hypothesis testing using the procedure of Benjamini et al. (2006), applied per column (i.e., separately for each (sub)rating and each rating adjustment). The sample period is from November 2012 to December 2020, but the exact starting date for each region is chosen so that there are at least 120 stocks in that region as of the starting date. Variable descriptions are provided in the Internet Appendix.

	No Adjustment				Industry-adjusted				Country-adjusted				Country & ind. adjusted			
	ESG (1)	E (2)	S (3)	G (4)	ESG (5)	E (6)	S (7)	G (8)	ESG (9)	E (10)	S (11)	G (12)	ESG (13)	E (14)	S (15)	G (16)
Asia-Pacific																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0	0/0	0/0	1/0	0/0	2/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/0	0/0
Emerging Countries																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
No. raters	9	8	8	8	9	8	8	8	9	8	8	8	9	8	8	8
No. hypotheses per cell	27	24	24	24	27	24	24	24	27	24	24	24	27	24	24	24
No. hypotheses per column	810	720	720	720	810	720	720	720	810	720	720	720	810	720	720	720

Table IA.5 - continued

	No Adjustment				Industry-adjusted				Country-adjusted				Country & ind. adjusted			
	ESG (1)	E (2)	S (3)	G (4)	ESG (5)	E (6)	S (7)	G (8)	ESG (9)	E (10)	S (11)	G (12)	ESG (13)	E (14)	S (15)	G (16)
Europe																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/1
FF3	4/4	0/0	2/2	0/0	1/1	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	2/2
Carhart	1/1	0/0	1/1	0/0	0/0	0/0	0/0	0/0	1/1	0/0	0/0	0/0	1/1	0/0	0/0	2/2
FF5	3/3	0/0	2/2	0/0	1/1	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/1
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	2/2
Japan																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
North America																
No controls	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
CAPM	0/0	0/0	0/0	0/0	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0	3/3	0/0	4/4	0/0
FF3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Carhart	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
FF5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/1	0/0
FF5+MOM	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/1	0/0
No. raters	9	8	8	8	9	8	8	8	9	8	8	8	9	8	8	8
No. hypotheses per cell	27	24	24	24	27	24	24	24	27	24	24	24	27	24	24	24
No. hypotheses per column	810	720	720	720	810	720	720	720	810	720	720	720	810	720	720	720

Internet Appendix: References

Armingeon K, Engler S, Leemann L, Weisstanner D (2023) Comparative political data set 1960-2021, Working paper.

Bali T, Engle R, Murray S (2016) *Empirical asset pricing: The cross section of stock returns* (John Wiley & Sons).

Beaver W, McNichols M, Price R (2007) Delisting returns and their effect on accounting-based market anomalies. *Journal of Accounting and Economics* 43(2-3):341–368.

Bessembinder H, Chen TF, Choi G, Wei KJ (2019) Do global stocks outperform U.S. treasury bills?, available at SSRN 3415739.

Botero J, Djankov S, Porta RL, de Silanes FL, Shleifer A (2004) The regulation of labor. *Quarterly Journal of Economics* 119(4):1339–1382.

Chaieb I, Langlois H, Scaillet O (2021) Factors and risk premia in individual international stock returns. *Journal of Financial Economics* 141(2):669–692.

Dyck A, Lins KV, Roth L, Wagner HF (2019) Do institutional investors drive corporate social responsibility? International evidence. *Journal of Financial Economics* 131(3):693–714.

Edmans A, Pu D, Zhang C, Li L (2024) Employee satisfaction, labor market flexibility, and stock returns around the world. *Management Science* 70(7):4357–4380.

Fama E, French K (1992) The cross-section of expected stock returns. *Journal of Finance* 47(2):427–465.

Griffin J, Kelly P, Nardari F (2010) Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. *Review of Financial Studies* 23:3225–3277.

Iliev P, Welch I (2010) Reconciling estimates of the speed of adjustment of leverage ratios, available at SSRN 1542691.

Jensen TI, Kelly B, Pedersen LH (2023) Is there a replication crisis in finance? *Journal of Finance* 78(5):2465–2518.

Krueger P, Sautner Z, Tang DY, Zhong R (2024) The effects of mandatory ESG disclosure around the world. *Journal of Accounting Research* 62(5):1795–1847.

Pástor L, Stambaugh R, Taylor L (2021) Sustainable investing in equilibrium. *Journal of Financial Economics* 142(2):550–571.

Shumway T (1997) The delisting bias in CRSP data. *Journal of Finance* 52(1):327–340.

Welzel C (2013) *Freedom Rising: Human Empowerment and the Quest for Emancipation* (United Kingdom. Cambridge University).

Welzel C (2015) World Values Studies key aggregates, waves 1-6 (Welzel replication file). GESIS Datenarchiv, Köln. ZA7470 Datenfile Version 1.0.0, https://doi.org/10.4232/1.12407.