Illuminating the Relationship Between ESG and Performance

By Guido Giese, PhD, Zoltán Nagy, CFA®, and Abhishek Srivastav, PhD
As interest has grown in investments that consider companies’ resilience to long-term environmental, social, and governance (ESG) risks, financial professionals have grown more interested in the relationship between ESG and investment performance.

The E, S, and G pillars themselves suggest a starting point. In this analysis, we examine drivers of risk and return for each of the pillars over time, as well as the relationship between ESG risks and specific issues that are used in calculating a company’s ESG rating. We further examine the relationship between such risks and the weighting of pillar-specific indicators in three approaches for developing an ESG rating.

The analysis highlights the significance of an investor’s time horizon in assessing the impact of ESG risks on companies’ financial fundamentals and stock-price performance. It illustrates the relative importance of ESG issues in comparing risks that can cause stock prices to move quickly with those that can erode a company’s competitiveness over time. And it shows how the weighting of each pillar in an ESG rating can impact long-term financial performance.

The results may help financial professionals create investment strategies that integrate ESG concerns.

**SOME ESG ISSUES ARE LONGER TERM THAN OTHERS**

Our analysis explores the relative impact of E, S, and G issues on companies’ stock-price performance during a 15-year period that ended in December 2021. Governance consistently showed more significance than environmental and social issues on financial fundamentals at any given point in time. Environmental and social issues’ contributions to stock-price performance, however, unfolded largely during the full study period.

The analysis suggests that markets may have more quickly priced in the risk of ESG events that materialized over the near term. Whereas governance has tended to encompass issues such as ethics breaches that are directly associated with near-term risks, a more limited set of environmental and social issues has been associated with such events. In contrast, issues such as carbon emissions are more likely to have eroded competitiveness and financial performance over a longer time horizon.

**EROSION RISK: LOOKING BEYOND ONE YEAR**

We start by analyzing stock—market performance during the full 15-year period of our study, on the assumption that some financial effects of companies’ aggregate ESG profiles may have unfolded slowly over time. To examine this effect, we created equal-weighted quintiles of top- and bottom-rated companies (according to their ESG scores) from the MSCI World Index based on company MSCI ESG ratings and each E, S, and G pillar score (standardized by sector). The quintiles were rebalanced monthly from December 2006 through December 2021. Figure 1 shows the cumulative monthly performance differential between the top-quintile (Q5) and bottom-quintile (Q1) companies, as reflected in a long—short strategy that buys the top-quintile portfolio and shorts the bottom-quintile one.
The Q5 quintiles all outperformed their Q1 quintile counterparts during our study period. The highest-scoring companies outperformed the bottom-scoring companies by between 13 percent (on E scores) and 39 percent (on cumulative ESG scores). The total ESG score—which is constructed from industry-specific weightings of the E, S, and G scores—exceeded each individual pillar’s score and was also the least cyclical.

The differences in time horizon mattered. The explanatory power of the E and G pillars, although significant, has waned in the past two years as S scores gained importance in explaining performance differences; this may indicate the importance of social factors for resiliency amid the pandemic. The data further shows that some effects of companies’ aggregate ESG profiles have unfolded slowly over time.

**EVENT RISK: LOOKING AT THE SHORT TERM**

We hypothesized that investors focus most intensely on events that could immediately affect company valuations. This effect tended to be captured more by issues such as fraud or corruption, which tie to the G pillar. The E and S scores, in contrast, comprised industry-specific environmental and social issues, only some of which could trigger tangible, event-driven risks such as accidents, strikes, or oil spills—and only for certain sectors. Other issues—such as managing human capital or carbon emissions, which carried high weights in the E and S scores of specific sectors—did not tend to erupt into tangible, negative events.

To probe this further, we used the frequency of stock-specific drawdowns as a measure for event-driven risk. For each month of the study period, we counted the number of companies that suffered drops in market value exceeding a given level during the following three years and compared the frequency of these drops for the highest- and lowest-scoring quintiles. Again, we used size-adjusted quintiles to ensure that potential differences in risk were not due to differences in size. Figure 2 shows the ratio of drawdowns observed in the top and bottom quintiles.

Compared with companies in the bottom-scoring quintiles, those in the top-scoring quintiles experienced fewer drawdowns. The finding held for each of the E, S, and G scores, with event-driven losses ranging from 50 percent to more than 90 percent of market value. For example, companies scoring in the lowest quintile on governance issues in a given month were, on average, 2.4 times more likely than the highest-rated companies on governance to lose more than 90 percent of their market value in the three years that ensued.

Among the three pillars, governance showed the biggest difference between top- and bottom-scoring companies in drawdown frequencies, followed by the social and environmental scores, respectively. That difference supports our hypothesis that governance-related incidents such as ethics breaches impacted a stock price quickly, with the greatest differences occurring for companies at the tail end of the loss threshold—i.e., with losses of more than 90 percent of market value.

The social score showed about half the difference as governance in the ratio of drawdown frequencies for top- and bottom-scoring companies. That may be because the social score reflects key issues—indicators that underlie the pillar scores and are proxies for ESG characteristics—for certain sectors that also tied to event risks. Such risks included accidents, strikes, or labor conflicts that could have affected a company’s stock price in the immediate to near term. Other S-pillar risks such as talent scarcity surfaced more slowly.

For the E pillar, we found a consistently lower ratio of loss frequency. Although some environmental risks such as toxic spills also could be construed as event-driven, the key issues that underlie the environmental score for certain sectors include issues such as carbon-emissions management and regulatory shifts that were less event-driven but may have affected companies’ businesses over longer periods.

**EXAMINING EVENT AND EROSION RISKS IN DEPTH**

To further illuminate implications of E, S, and G ratings for companies’ competitiveness over both near- and long-term horizons, we examined the impact of each ratings pillar on key issues in the context of both event and erosion risks during the 15-year period ended December 2021 (see figure 3).

Our analysis focused on the following 11 key issues that are most commonly used by MSCI ESG Research in calculating a company’s ESG rating:
Environmental (E pillar): carbon emissions, water stress, toxic emissions, and waste
Social (S pillar): labor management, health and safety, human-capital management, and privacy and data security
Governance (G pillar): Corporate governance, business ethics, corruption and instability, and anticompetitive practices

To quantify event and erosion risks, we used equal-weighted portfolios for each key issue and compared the Q5 quintile (best ESG characteristics) to the Q1 quintile (worst ESG characteristics) using the following measures:

**Event risk.** As above, we used the Q1/Q5 frequency ratio at a 90-percent loss level as a measure of event risk (shown along the y-axis in figure 3); the higher the ratio, the more effective the pillar as an indicator.

**Erosion risk.** To measure erosion risk (shown along the x-axis in figure 3), we compared the annualized cumulative stock-specific contribution to the relative performance of top-scoring (Q5) versus bottom-scoring (Q1) companies for each key issue. By examining stock-specific returns only, we controlled for the effect of other known factors on performance.\(^8\)

We found that nine out of 11 issues showed a positive Q5–Q1 stock-specific performance contribution cumulatively during the study period, and six of the 11 issues showed a positive contribution in terms of identifying differences in event-driven risks. We also saw distinct differences across the issues categorized under the E, S, and G pillars.

**ENVIRONMENTAL KEY ISSUES**
Of the three environmental key issues, carbon emissions and toxic emissions were driven by erosion risk. They showed positive long-term differences between the top- and bottom-scoring companies but did not mitigate event risk. Though water stress, the third environmental issue, did not display either economically meaningful erosion or event-driven risk during our study period, constraints on water appear to be emerging as an important factor in some sectors and, consequently, could impact companies in water-stressed countries.\(^9\)

**SOCIAL KEY ISSUES**
The key issues under the social pillar showed more balanced results, with differences in labor management, e.g., mitigating labor conflicts, showing strong event- and erosion-risk characteristics. In fact, companies with top scores in labor management outperformed bottom-scoring companies by an average of 1.08 percent in residual stock-specific return per year and showed significant reductions in event risks (the top-scoring companies experienced severe stock-price losses one-third as frequently as the low scorers). The human capital key issue aligned with both event- and erosion-driven risk. Although health and safety key issues showed strong long-term idiosyncratic performance differences, such issues showed minimal differentiation for event risk between high- and low-scoring companies. Though high-profile data breaches such as those at Equifax and Facebook (now Meta) might have suggested otherwise, differences in companies’ management of privacy and data security did not historically contribute to positive performance during the study period. Nor did top-scoring companies on those key social issues avoid more negative events than low-scoring companies.

**GOVERNANCE KEY ISSUES**
Governance-related key issues as a whole showed the strongest results along both risk dimensions during our study period. The key issues of business ethics and anticompetitive practices showed much stronger event-risk characteristics (and minimal erosion risk). The bottom-scoring companies on business ethics were about four times more likely than top-scoring companies to experience a steep drop in the price of their stock during the three years that ensued; the bottom-scoring companies on corruption were more than twice as likely as the top-scoring companies to experience a sharp falloff.

Corporate governance, in contrast, and especially the corruption key issue, correlated positively with long-term...
differences in financial performance, with stronger erosion-risk characteristics and less differentiation based on event risk.

LOOKING AT ESG KEY ISSUES OVER THE LONG TERM

Erosion-driven ESG issues, of course, grab fewer headlines than the more abrupt (and sometimes dramatic) shifts that event-driven risks can trigger. So we also examined how ESG key issues performed during the entire 15 years of our study period, starting with the aggregate ESG score.

As previously, we focused on understanding performance differentials arising from common factors as well as firm-specific returns; differences in performance likely reflect common factors as well. Table 1 further contextualizes the long-term performance of key issues. It shows the stock-specific performance of the equal-weighted Q5 and Q1 portfolios during the sample period together with the contributions of the main equity factor groups to it. Companies in the top quintile of MSCI ESG scores, for example, outperformed those in the bottom quintile by 2.5 percent annually in absolute terms and by 1.1 percent annually in terms of stock-specific performance contribution.

Companies in the top quintile outperformed those in the bottom quintile for eight of 11 key issues that we tested. The stock-specific performance shows a similar picture: Nine of 11 key issues showed a positive stock-specific contribution. On average, key issues for the G pillar showed the highest stock-specific performance (0.9 percent) followed closely by the E pillar (0.7 percent). Among factors, style factors, which ranged from −1.7 percent to 3.5 percent annually, showed the most significant contribution. Stock-specific returns ranged between −0.5 percent and 1.5 percent, indicating that the “pure” residual effect of key issues over the long term could easily be overshadowed by other unintended systematic effects.

HOW THE WEIGHTING OF ESG RATINGS AFFECTED PERFORMANCE

Because the analysis suggested that environmental and social issues were more industry-specific and tended to show up in financial measures over a longer time frame than governance issues, we also sought to assess implications for investors in combining E, S, and G issues into an aggregate ESG score or rating.

To do so, we investigated three different approaches to creating a combined ESG score or rating: equal weighting; an optimized approach that sets weights based on historical data; and industry-specific weights that vary over time as represented by MSCI’s ESG Ratings.

APPROACH #1: EQUAL WEIGHTS

Equal weighting (or so-called naive weighting) offers the benefit of simplicity, transparency, and comparability across industries. If an investor does not have specific views about the relative importance of E, S, or G issues but wants to take advantage of information in ESG data, an equal weighting could be appropriate.

For equal weighting, we computed a monthly aggregate ESG score for each company between December 2006 and December 2021 that comprised one-third E key issue scores, one-third S key issue scores, and one-third G key issue scores.

APPROACH #2: OPTIMIZED WEIGHTS

We also considered an optimized weighting based on historical data. An optimized approach may help investors who do not have a specific view and would rather the data itself dictate the optimal E, S, and G weights based on their historical significance.

To create an optimized ESG score, we chose company fundamental data to

Table 1

<table>
<thead>
<tr>
<th>ESG score</th>
<th>Active</th>
<th>Styles</th>
<th>Industries</th>
<th>Countries</th>
<th>Stock-specific</th>
<th>Currencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESG score</td>
<td>2.5%</td>
<td>2.6%</td>
<td>0.2%</td>
<td>0.0%</td>
<td>1.1%</td>
<td>−1.4%</td>
</tr>
<tr>
<td>Carbon emissions</td>
<td>2.5%</td>
<td>2.1%</td>
<td>1.2%</td>
<td>−0.2%</td>
<td>0.7%</td>
<td>−1.2%</td>
</tr>
<tr>
<td>Water stress</td>
<td>0.4%</td>
<td>2.3%</td>
<td>−0.1%</td>
<td>−0.3%</td>
<td>−0.1%</td>
<td>−1.4%</td>
</tr>
<tr>
<td>Toxic emissions</td>
<td>0.3%</td>
<td>1.5%</td>
<td>−1.2%</td>
<td>−0.4%</td>
<td>1.5%</td>
<td>−1.1%</td>
</tr>
<tr>
<td>Labor management</td>
<td>2.0%</td>
<td>1.1%</td>
<td>0.1%</td>
<td>0.3%</td>
<td>1.1%</td>
<td>−0.6%</td>
</tr>
<tr>
<td>Health &amp; safety</td>
<td>2.3%</td>
<td>3.5%</td>
<td>0.0%</td>
<td>0.2%</td>
<td>0.3%</td>
<td>−1.7%</td>
</tr>
<tr>
<td>Human capital</td>
<td>−0.4%</td>
<td>0.8%</td>
<td>0.2%</td>
<td>−0.9%</td>
<td>0.1%</td>
<td>−0.6%</td>
</tr>
<tr>
<td>Privacy &amp; data security</td>
<td>−0.2%</td>
<td>0.9%</td>
<td>0.2%</td>
<td>−0.5%</td>
<td>−0.5%</td>
<td>−0.3%</td>
</tr>
<tr>
<td>Corporate governance</td>
<td>2.7%</td>
<td>2.3%</td>
<td>−0.7%</td>
<td>0.8%</td>
<td>0.9%</td>
<td>−0.6%</td>
</tr>
<tr>
<td>Business ethics</td>
<td>0.9%</td>
<td>0.1%</td>
<td>−0.7%</td>
<td>0.4%</td>
<td>0.8%</td>
<td>0.3%</td>
</tr>
<tr>
<td>Corruption</td>
<td>3.6%</td>
<td>3.3%</td>
<td>−0.8%</td>
<td>0.9%</td>
<td>1.2%</td>
<td>−1.1%</td>
</tr>
<tr>
<td>Anticomp practices</td>
<td>−2.5%</td>
<td>−1.7%</td>
<td>−1.6%</td>
<td>0.2%</td>
<td>0.6%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Stock-specific performance, highlighted in green, shows the relative annualized performance, net of systematic effects, of the highest- and lowest-rated portfolios.

Source: MSCI ESG Research, based on the 15-year period ended December 31, 2021. Analysis uses GEMLT for return attribution. Active return in USD.
represent investor objectives. Specifically, we looked for a combination of E, S, and G pillar scores that maximized the economic effect via three transmission channels (cash-flow, idiosyncratic risk, and systemic risk) using gross profitability, residual volatility, and systematic volatility, respectively, as the target financial variables. We limited the sample period to the one that preceded the pandemic to avoid the shocks to market valuations that occurred in the year that ensued.

Note that although the sample period was between December 2006 and December 2019, we also applied the optimal combination out-of-sample for the two years that ended December 2021. That allowed us to determine whether differences in approaches to weighting ESG ratings might illuminate how equities recovered as the pandemic ebbed. This approach resulted in the following weights: 25 percent for E, 5 percent for S, and 70 percent for G.

**APPROACH #3: INDUSTRY-SPECIFIC WEIGHTS**

The third approach to selecting and weighting E, S, and G issues for each industry (the approach used in creating MSCI’s ESG Ratings) more precisely reflects industry exposures to E, S, and G risks. It has the drawback, however, of introducing complexity and limiting comparisons among industries.

On average, MSCI’s ESG Ratings reflect six key issues in each of 158 Global Industry Classification Standard (GICS®) sub-industries. The selection of key issues and their respective weights are readjusted annually through a process that combines quantitative assessment of industry exposures to emerging issues and consultations with investment practitioners.

Using this process, weights have varied over time across sectors. During our 15-year backtest, there were more than 2,000 possible permutations of E, S, and G weights. As of the end of 2021, the weight of the E pillar ranged from 5.3 percent for the communication services sector to 47.0 percent for utilities; the weight of the S pillar ranged from 17.5 percent for utilities to 53.1 percent for the health care sector; and G score weights varied between 33 percent for materials and 44 percent for the industrials sector.

During the backtest period, the pillar weights averaged 30 percent for E key issues, 39 percent for S key issues, and 31 percent for G key issues. These weights showed significant variation over time. The average G pillar weight increased from 18 percent in the first half of the sample period (2007–2013) to 29 percent in the second half (2014–2021), highlighting the increasing importance that MSCI ESG Research has assigned to governance issues over time.

**LONG-TERM FINANCIAL SIGNIFICANCE**

During the full 15-year time horizon of our study, the industry-specific approach represented by MSCI’s ESG Rating outperformed both the equal-weighted ESG score and the backtest-weighted ESG score by 5 percent and 14.6 percent, respectively.

Figure 4 shows the stock–price performance difference between the top-quintile companies and the bottom-quintile companies for each of the three approaches. We found that the industry-specific weighted approach represented by the overall MSCI ESG Rating approach correlated to better stock performance during the study period and showed a lower level of cyclical.

Note that the optimized score’s out-of-sample performance after 2019 was weak compared with either the equal-weighted or MSCI-weighted ESG scores. This may alert investors to the limits inherent both in relying too heavily on historical data in ratings construction and the potential pitfalls of data mining.

When looking at long-term financial significance, we found that social and environmental key issues, which tended to unfold more slowly, became more important over time. Yet a long-term view, by itself, may not reveal the full story. The equal-weighted ESG score had nearly the same average weight distribution to E, S, and G as the MSCI ESG score.

A key difference is the dynamic nature of the MSCI ESG score. For its ratings, MSCI adjusted both the selection of ESG issues and the weighting for each of the 158 GICS® sub–industries annually. The shifting balance between E, S, and G key issues might help explain the superior financial performance of that approach compared with static weighting schemes over the study period.
We used the frequency of stock-specific drawdowns as a measure for event-driven risks to probe performance differences among ESG metrics. Here, too, for each month of the study period we counted the number of companies that suffered drops in market value beyond a given threshold during the ensuing three years and compared the frequency of these drops for the highest- and lowest-scoring size-adjusted quintiles. Figure 5 shows the ratio of drawdowns observed in the top and bottom quintiles.

Although companies in the top-scoring quintile of ESG scores experienced fewer drawdowns than those in the bottom-scoring quintile, the relative impact varied depending on how we computed the ESG score. The top quintile of companies scored with an industry-specific weighting were three times less likely to lose 90 percent of their market value compared with companies in the bottom-scoring quintile. The ratio fell to 2.2 for companies in the top-quintile scored with an equal-weighted score, and to two for those scored using an optimized weighting.

COMBINING INDICATORS
Weighting schemes can play an important role in fine-tuning ESG-rating methodologies by enhancing their forward-looking assessment of ESG risks and how such risks may be reflected in the rating model, the analysis shows.

In the short term, we found that both equal-weighted and optimized approaches were more heavily weighted toward governance issues, but that short-term correlations did not mean long-term financial significance. The reverse was true for an approach that adjusted the weights of E, S, and G key issues dynamically by industry; the approach displayed strong financial performance over the long term at the expense of short-term correlations to key financial variables.

An optimization-based approach that used historical data and a static target function was too simplistic and too backward-looking. The risks that investors face are anything but static.

Investors aiming to integrate ESG factors to achieve better long-term financial results often have overlooked the significance of how ESG indicators combine. In our study, the weighting scheme that achieved the strongest significance over one year showed the weakest significance—as measured by stock-price performance—during the 15 years we examined.

IMPLICATIONS FOR INVESTORS
Our findings hold important implications for ESG integration in portfolio construction. Investors building concentrated stock portfolios with relatively high turnover may want to focus on identifying and mitigating short-term event risks. Such investors may find erosion risks to be less relevant given their tendency to unfold over longer time horizons.

Investors building broadly diversified portfolios with long investment horizons, including, for example, indexed or buy-and-hold investors, may be more focused on long-term erosion risks in their choice of ESG criteria and integration. They also may aim to mitigate event risks through diversification. Such investors could still face headline risk from unexpected events but may regard it as a reasonable trade-off compared with the potential for erosion risks.

How investors aggregate the pillars of ESG also can hold value. Our findings suggest that an approach to constructing ratings that adjusts industry-relevant issues and weights annually can make a difference to financial performance over time compared with an approach that assembles the E, S, and G components arbitrarily or based chiefly on historical data.

Guido Giese, PhD, is global head of ESG and climate solutions research at MSCI, where he leads applied research and thought leadership on ESG integration, impact investing, and climate investing. He earned a PhD in applied mathematics from the Swiss Federal Institute of Technology Zurich. Contact him at guidogiese@msci.com.

Zoltán Nagy, CFA®, is a member of the MSCI research team focused on ESG and climate solutions. He specializes in the integration of climate and ESG considerations into equity portfolio management. He earned a PhD in theoretical physics from the University of Cergy-Pontoise and an engineering degree from the École Polytechnique. Contact him at zoltan.nagy@msci.com.

Abhishek Srivastav, PhD, was a vice president at MSCI, where he focused on conducting applied research across various ESG topics, including climate change, executive compensation, and banking. He earned a PhD in finance from the University of Edinburgh. Contact him at abhishek.1827@gmail.com.
Russia’s invasion of Ukraine is likely to increase risk for investors by requiring a faster transition to a low-carbon economy in the long term after an initial delay, an analysis by MSCI ESG Research shows.*

Delaying the transition to net zero by emitting more greenhouse gases from swapping coal for the burning of Russian gas in Europe could raise net downside risk by 17 times on average across 11 sectors, according to the analysis, which compared a climate trajectory for companies in the MSCI Europe Investable Market Index (IMI) that aims to keep the rise in average global temperatures well below 2°C Celsius with a scenario in which emissions continue to rise and decarbonization efforts are delayed by up to a decade.

To limit the rise in average global temperatures to 1.5°C—the threshold for averting the worst impacts of a warming planet—global greenhouse gas (GHG) emissions must remain within a “budget” of 400 gigatons. (A gigaton equals one billion tons.)

But more GHG emissions in the short term would have to mean fewer emissions later if Europe were to meet its climate goals (see figure S1). “In the long term, delaying the shift toward cleaner sources of energy while remaining within a net-zero global emissions budget would require an accelerated shift later, according to science-based climate scenarios,” write MSCI’s Chris Cole, Zoltán Nagy, and Guido Giese, who used MSCI’s Climate Value-at-Risk (Climate VaR) metric to compare the financial risk for constituents of the MSCI Europe IMI for the two hypothetical scenarios.

Differences in net transition risk—the sum of downside risk from incurring a high carbon price and the upside potential from an expanded market for clean-tech infrastructure—between a delayed transition and a below 2°C scenario were largest for the energy and materials sectors, with downside risk of 60 percent versus 10 percent and 59 percent versus 11 percent, respectively (see figure S2). Net transition risk for the communications-services, consumer-staples, and consumer-discretionary sectors led to financial losses roughly 24, 17, and 12 times higher, respectively, in the two scenarios.

The utilities and energy sectors had potentially large technology opportunities in a delayed transition (66 percent and 34 percent, respectively), though higher policy risk outweighed the opportunities (79 percent and 93 percent, respectively), for net downside risk.

The analysis looked specially at the MSCI Europe IMI; however, an analogous analysis of the MSCI USA IMI found similar results between the two scenarios.

ENDNOTES

1. ESG ratings vary in the information they are designed to measure. Our analysis focuses on ESG ratings that measure financially relevant environmental, social, and governance risks.


3. The MSCI World Index captures larger and mid-cap representation across 23 developed-market countries. With 1,540 constituents, the index covers approximately 85 percent of the free float-adjusted market capitalization within each country, as of May 31, 2022.


5. The ratio hovered just above one.

6. For example, airlines face mandatory requirements to further reduce their fleets’ carbon intensity by 2026 through the Carbon Offsetting and Reduction Scheme for International Aviation (CORISA).

7. Key issues are selected and weighted for each Global Industry Classification Standard (GICS®) sub-industry based on potential exposure to the respective issue. The GICS® was jointly developed by MSCI and S&P Global Market Intelligence. For drawdowns, our sample period effectively ends in 2018 because we calculate drawdowns based on returns over the subsequent 36 months.

8. We used the MSCI Global Total Equity Market Model for Long-Term investors (GEMLT) to understand the portion of the observed performance differentials that can be attributed to common factor exposures as well as the residual performance effect. We refer to this as stock-specific or idiosyncratic performance.


10. See footnote 7, infra.