Best Practices in ESG Investing

Herb Blank, Gregg Sgambati, and Zack Truelson

Herb Blank
is the head of business development at S-Network Global Indexes, Inc., in New York, NY.
hblank@snetworkinc.com

Gregg Sgambati
is the head of ESG Solutions at S-Network Global Indexes, Inc., in New York, NY.
gsgambati@snetworkinc.com

Zack Truelson
is a marketing associate at S-Network Global Indexes, Inc., in New York, NY.
ztruelson@snetworkinc.com

Environmental, social, and corporate governance (ESG) investing has exploded. As of January 2015, investing in ESG portfolios had increased 76% since 2012 and accounted for more than one-sixth of all investing (McQuillen [2015]). As the more conscientious approaches to investing have evolved from simple avoidance of "sin stocks" (such as those in tobacco and alcohol) to increasingly nuanced and performance-driven strategies, demand has grown for reliable data. While popular interest in such strategies has been on the rise for decades, their focus and terminology have shifted away from "ethical investing" and toward "socially responsible investing" (SRI) and similar names (Eccles and Viviers [2011]). From the beginning, the trend has been somewhere under the collective umbrella of ESG issues, but recently, we have seen attention more broadly distributed to cover the full scope of ESG. Instead of using such a blunt instrument as negative screening at the industry level, more investors are paying attention to what distinguishes companies *within* industries, opening an array of new inquiries about how—rather than just what—the companies produce.

ESG/SRI-oriented methods have been around for several decades, dating back at least to the launch of the Calvert Social Investment Fund in 1982. It was not always as data-driven as it is today. Although it is not a random process as operationalized by Adler and Kritzman [2008]—who used Monte Carlo simulations to argue that SRI incurs opportunity cost—screening that excludes and includes companies based solely on their industries and products cannot be expected to work as a reliably profitable strategy. According to a thorough analysis by Renneboog, Ter Horst, and Zhang [2008a], as of December 2003, investors were in transition from simple negative screening to more ESG-oriented methods. Globally, the screen ratios were mixed, but particularly in the United States, negative screens (97%) dominated positive (69%) and anti-"sin" screens (92%) outweighed environmental (72%) and social/governance (68%), with ethical screens comparable at 57%. (These are the percentages of funds that employed each type, so they sum to more than 100%.) And not surprisingly, results were likewise mediocre: another analysis (by the same authors) of socially responsible (SR) funds from 1992 to 2003 found them to underperform (Renneboog, Ter Horst, and Zhang [2008b]).

More recent findings are still mixed. Humphrey and Tan [2013] found no effect on returns from either positive or negative screens, while Trinks and Scholtens [2015] found a significant opportunity cost for negative screening. Statman and Glushkov [2008] found that negative screening of "shunned"

stocks (those involved with tobacco, gambling, alcohol, weaponry, and nuclear power) led to lower returns while selecting for higher ESG performance (using KLD scores) led to higher returns. They argued that these contrary effects largely canceled out in the world of SRI funds, resulting in no net benefit. But their study does suggest that a focus on ESG performance as opposed to negative screening holds promise in terms of performance.

Much depends on how the researchers choose to model the behavior of the portfolio manager. The study by Adler and Kritzman modeled the stock selection as a randomized negative screen, essentially excluding a given percentage of the universe of stocks without regard to any of the stocks' attributes. De and Clayman [2015] also used a method of randomized portfolio creation, but did so after restricting the universe of stocks according to Thomson Reuters/ASSET4 ESG ratings. They found that in real-world practice, there is clearly room for improvement. Wimmer [2013] found that the average ESG ratings of the portfolios of SR mutual funds persist for only two to three years—but rather than being caused by high volatility in the ESG ratings themselves, the short time horizon is caused by turnover in the portfolios. In other words, it is not that highly rated companies deteriorate and earn lower ESG ratings after such a short time, but rather that the managers of the SR funds are swapping out companies apparently with little regard for what the ESG ratings are telling them. We hope that as ESG ratings and similarly informative data gain prominence, they will become more wisely used.

But in order to address those new inquiries on a large scale, we need objective measures. Negative screening of objectionable products requires very little research; either a company produces a product that a particular investor finds objectionable, or it does not. Additionally, such basic product information is the kind that companies invariably provide. But when we start asking questions that could distinguish between two companies making the same product, we soon run into obscurity, opacity, and subjectivity where we need clarity, transparency, and objectivity.

Simply put, comparable ratings enable better comparisons. They allow researchers to study the statistical influence of ESG practices, and they allow investors to apply the researchers' findings in ever-more-refined strategies. They encourage companies to be more transparent in their practices, and they have the potential to encourage consumers to spend with more attention to ethics and sustainability.

AN OBJECTIVE STANDARD

The Thomson Reuters corporate responsibility ratings (TRCRR) are designed to serve as an objective standard for assessing and comparing companies' social responsibility and are joined by a number of similarly conceived rating systems.[1] These ratings provide a quantitative grounding for the purposes of both research and investing, whether aiming at the whole of ESG or at the individual ESG components (which are referred to as "pillars" in the TRCRR system).

As detailed by Blank [2013], the development of the TRCRR involved hundreds of primary researchers and a lot of careful thought, resulting in different designs for the three ESG pillars. The goal was to take a vast array of information and distil it into numerical ratings that could be used to fairly compare companies across and within industries and geographical regions. Achieving that goal required numerous steps of converting, weighting, and normalizing scores, as well as qualitative decisions about how to group companies and how to weigh the importance of various pieces of data.

In this article, we will describe the path of refinement from the raw data points to the final scores, based on the complete methodology detailed in the TRCRR Rule Book [2014]. To make the description as clear as possible, we will follow three examples as they evolve from qualitative to quantitative, subjective to objective, raw to normalized. You will gain an understanding of how the ratings address the challenges we have discussed. Although you will see a number of equations provided for clarity's sake, the procedures they represent will also be explained verbally, so you will gain an intuitive grasp of the information, even without having to parse mathematical expressions.

ASSEMBLING THE RAW DATA

The most labor-intensive part of the process is at the beginning, when Thomson Reuters ASSET4 (henceforth ASSET4) researchers collect more than 500 data points for each of some 5,000 companies. A typical company's profile takes a week to assemble, drawing from a variety of sources including stock exchange filings, annual reports, and news outlets. These data points are sorted at two levels, first into eighteen categories,

second into three pillars—environmental, social, and corporate governance.

We will follow one example from each of the three pillars. The first, from the environmental pillar, asks, "Is the company under the spotlight of the media because of a controversy linked to biodiversity?"

We should note a couple of things about this question and hold them in our memory for the next section, in which we will see how the raw data are distilled into more than 250 key performance indicators (KPIs). First, note that this question requires a "yes" or "no" answer, which makes it a *Boolean* rather than a *metric* question (which might be answered either on a 1–5 scale or with an absolute number). Second, note that answering "yes" is not a good thing in this case. That means that this question is one of *negative polarity* (if a "yes" answer were a good thing, then the question would have *positive polarity*).

The second example is from the social pillar: "Does the company report or show to use human rights criteria in the selection or monitoring process of its suppliers or sourcing partners? And does the company report or show to be ready to end a partnership with a sourcing partner if human rights criteria are not met?"

Note that this example actually consists of two distinct questions, but they will be lumped together into a single KPI because the second question is dependent on the first; a company that does not pay attention to human rights when selecting its suppliers is not likely to drop those suppliers for human rights-related reasons. But we would not want to discard the second question, because follow-through is an important factor to measure. (In the next section, we will see how partial credit is given for "yes/no" and "no/yes" answers to these dual-question KPIs.) Also note that both of the questions are of positive polarity, unlike that in the first example.

The third example is from the corporate governance pillar, and it requires simply the "percentage of shares held by all insiders and 5% owners." Hopefully, you recognize this as a metric question, as opposed to a Boolean one. This metric question in particular is asking for an absolute number rather than a number on the relative 1–5 scale, and converting so many disparate quantities into comparable ones will bring us to the first of many equations in this article (but accompanied by accessible wording, never fear!). And thankfully, due to the way the conversion works, we do not need to be concerned whether this question is of positive or negative polarity.

KEY PERFORMANCE INDICATORS

The KPIs can be thought of as the basic units in creating the ratings. They represent the first stage in which the data points are roughly comparable, all converted to values between zero and one. The KPIs will be the units to which we apply various weights and assign relative levels of importance (RLIs), and which we group together by industry, region, both, or neither. To offer an analogy from linguistics, a KPI is to a morpheme (the smallest meaningful unit of speech) as an individual data point is to a phoneme (a mere sound).

Boolean KPIs are created using one of the following conversion tables (Exhibit 1 being for positive polarity, and Exhibit 2 being for negative polarity):

Answering "yes" to a one-question KPI is scored the same as answering "yes" to both questions in a two-question KPI. If the company answers "yes" to one and "no" to the other, the score is 0.5. Note how "not relevant" (NR) and "not available" (NA) are treated differently. "Not relevant" means that the question is irrelevant

EXHIBIT 1
Positive Polarity KPI

Response	Value
Yes	1
Yes/Yes	1
Yes/No	0.5
No/Yes	0.5
N/R (not relevant)	0.5
No/No	0
No	0
NA (not available)	0

EXHIBIT 2
Negative Polarity KPI

Response	Value
Yes	0
Yes/Yes	0
Yes/No	0.5
No/Yes	0.5
N/R (not relevant)	0.5
No/No	1
No	1
NA (not available)	0

to the company's region, industry, or both. For example, the question "Is the company developing hybrid vehicles" is relevant to the automobile industry, but decidedly not to the telecommunications industry. A car company that invests nothing in developing hybrid vehicles may be fairly docked in its ESG rating, but it would be unfair to similarly dock a telecom company, because a typical company in that sector could not be expected to have activity in vehicle production of any kind.

"Not available" is treated more stringently, as this means that the company is simply not providing the information. We score this lack of an answer the same as the lowest-scoring answer to that question given by any comparable company (that is, one in the same region, industry, or both). The logic behind this method is as follows: If we scored NA as 0.5 like we do with NR, then companies would have an easy way to game the ratings. For any question to which their answer would be negative, the company would be incentivized to refuse to give any answer, thereby getting a 0.5 instead of a 0. By scoring NA as the worst comparable score, we incentivize companies to provide the information, but we also do not punish them unduly for lacking it, as they can never be the worst-scoring company as a result.

We can easily see where our environmental and social pillar examples would fall in the previous exhibits. Although that is quite straightforward, the metric KPIs require some calculation to make the quantities comparable. Our corporate governance example will undergo the following procedure:

1. Subtract the lowest reported number (among regional peers) from the company's reported number.
2. Subtract the lowest reported number from the highest reported number (again, that's the highest reported from among companies within the same geographical region).
3. Divide the result of (1) by the result of (2).
4. Subtract the result of (3) from one.

As you might expect, the above steps will vary depending on the polarity of the question—if our metric example had been of positive polarity, then we would have skipped (4). But there is also a twist for metric KPIs in the environmental pillar, which undergo two additional steps: the number—that is the quotient from (3), or if applicable, the difference from (4)—is multiplied by 0.4, and then added to 0.6. This means that an environmental KPI score cannot be lower than 0.6, which is higher than the score given for NR (0.5). The score given to an environmental KPI for NA is 0.4.

Alternatively, we can express the preceding in an equation (this one being for a negative polarity environmental metric KPI):

$$Score = 0.6 + 0.4 * \left(1 - \left(\frac{x_i - \min\limits_{x \in I} x}{\max\limits_{x \in I} x - \min\limits_{x \in I} x} \right) \right) \qquad (1)$$

where $x \in I$ tells us that when we are finding the maximum reported value x, we are considering only the values reported by companies within the same industry I. The symbol $\in$ means "is an element of," so the expression as a whole is telling us that we are taking the maximum value of x *given that* x is reported by a company in the same industry as the company that reported x_i, the individual value we are converting into a KPI score.

Why do we use the extra steps for environmental metric KPIs? Many KPIs in that pillar have low rates of reporting, which would result in numerous scores of zero. To adjust for that and preserve comparability across pillars, we set the NA score for metric KPIs at 0.4 instead of 0 and use the additional steps for reported metrics, so that reporting anything guarantees a company a score of higher than NA or NR, starting at 0.6 and maxing out at 1. The NA scores being set at 0.4 gives the distribution a more normal shape (recall that the NR score is 0.5, halfway between the NA score and the minimum score for a reported metric).

For the positive polarity version, we would simply remove the subtraction from one, and for an analogous KPI in the corporate governance or social pillars, we would furthermore remove the multiplication by 0.4 and the addition to 0.6. To prevent any confusion, we have provided a full list of the equations in the Appendix.

RELATIVE LEVELS OF IMPORTANCE

The second step is to assign each KPI an RLI. RLIs range from 0 to 5 and are assigned based on several factors:

1. How relevant that KPI is to the peer group (that is, companies within the same industry in the

environmental pillar; the same region in the corporate governance pillar; and in the social pillar, the same industry for some KPIs, the same region for others, and the entire universe for a few).
2. The percentage of companies in the peer group that provide answers for the KPI.
3. The range, skewness, and standard deviation of answers to the KPI.
4. The KPI's independent information content.
5. How objectively measurable the KPI is.
6. How robustly the statistical results for the KPI can be validated using the academic literature.

As you can see, assigning the RLIs involves some degree of subjective judgment and a great deal of simplification, as all these considerations condense into a single value. That said, there are no subjective judgments being made on a company-by-company basis; all of the above considerations translate into quantitative indicators (if they do not begin as such) and are applied consistently, so we need not think of the RLI assignment as any more hazardous than, for example, the two additional mathematical steps used to convert metric KPIs in the environmental pillar. Any arbitrariness is spread evenly across companies, so none is especially advantaged or disadvantaged.

CALCULATING THE WEIGHTS

Once assigned, each RLI is modified by a multiplier that is based on the percentage of companies in the peer group that gave an answer for the KPI in question. The multiplier can take one of three values: 0, 0.5, or 1. The percentage thresholds that determine the multiplier vary by pillar; the environmental pillar has lower reporting percentages in general compared with the other two, so its thresholds are adjusted downward.

- Weights in the Environmental Pillar: 0 (for when less than 10% of the companies in the peer group gave answers to the KPI), 0.5 (for when between 10% and 30% gave answers), or 1 (for when more than 30% gave answers).
- Weights in the Corporate Governance and Social Pillars: 0 (for when less than 0.5% of the companies in the peer group gave answers to the KPI), 0.5 (for when between 0.5% and 15% gave answers), or 1 (for when more than 15% gave answers).

Next, KPI weights are determined. Uniquely in the environmental pillar, they are split by ten factors. Because the environmental pillar in particular has low reporting rates, it is especially sensitive to the comparability within the peer group (in this case the company's industry). It turned out that reporting percentages could be grouped more cleanly by the factor of the KPI than by the industry of the company. Since the divisor weight is redistributed within the industry, the divisor needs to differ by factor to adjust for the differences in reporting across industries and ensure proper peer-to-peer comparisons.

Most of the factors map onto one within-pillar category each, except for two of the factors (7 and 10) that are represented in all three categories of the environmental pillar (resource reduction, emission reduction, and product innovation). Exhibit 3 shows all the factors.

To calculate an environmental KPI's weight, we add together all the modified RLIs of the KPIs within a factor to get a divisor for that factor, divide the KPI's RLI by the factor's divisor, and finally multiply by the factor's weight, as follows:

$$Environmental\ KPI\ Weight = \left(\frac{RLI}{\sum \left(RLI * Multiplier \right)} \right) * Factor\ Weight \qquad (2)$$

We then dynamically scale the scores on the KPI for the peer group, such that the best score is equal to 1 and the worst score is equal to 0.

For the social and governance pillars, we likewise add the RLIs for the KPIs in each peer group to create divisors for the respective peer groups and divide the

EXHIBIT 3
Ten Factors of the Environmental Pillar

Factor No.	Name	Type
1	Emission Reduction Policy Driver	Boolean
2	Product Innovation Policy Driver	Boolean
3	Resource Reduction Policy Driver	Boolean
4	Observable Emission Reduction Practices	Boolean
5	Observable Product Innovation Practices	Boolean
6	Observable Resource Reduction Practices	Boolean
7	Controversies	Boolean
8	Emission Reduction Tonnes/Revenues	Metrics
9	Resource Reduction Tonnes/Revenues	Metrics
10	Leadership	Metrics/Boolean

RLI by the peer group's divisor, but then go straight to the scaling step (there is no need to multiply by the factor's weight, since the factor divisions exist only in the environmental pillar).[2]

After replicating the above process for every KPI, we can calculate the raw score for each pillar simply by adding together all the products of the KPI values and their weights. In this way, all the environmental KPIs combine into an environmental raw score, and likewise with the social and corporate governance KPIs, thereby creating three raw pillar scores that we will convert into finalized ratings in the next section.

Every KPI is benchmarked according to industry, region, or the entire universe of companies. In other words, a company's score on a given KPI may alternatively be compared with the scores of other companies in the same industry or region, or simply all the companies in the universe. Corporate governance KPIs are all compared within regions; environmental KPIs are all compared within industries; and social KPIs are variously compared within regions, industries, and the universe.

FROM RAW SCORES TO RATINGS

This final section of methods contains the most steps and is perhaps the most abstract. It will take us from the raw pillar scores to the final, publishable pillar ratings (which are combined to form the composite ESG scores).

First, we calculate the *Z*-score for each raw score. The *Z*-score is a way of expressing how much the raw score deviates from the mean compared with the other raw scores. It is simply the number of standard deviations away from the mean that the raw score falls, with negative values meaning so far below the mean and positive values meaning so far above the mean, as follows:

$$Z\text{-}score = \left(Raw\ Score - \mu(Raw\ Scores)\right) / \sigma(Raw\ Scores) \tag{3}$$

Second, we adjust the distribution for outliers. If the *Z*-score is less than –3, then the adjusted *Z*-score is the lowest *Z*-score that is greater than or equal to –3. This is mirrored on the positive side: if the *Z*-score is greater than 3, then the adjusted *Z*-score is the highest *Z*-score that is less than or equal to 3. In other words, the *Z*-scores with absolute values greater than 3 are changed to equal those that have the greatest absolute values equal to or less than 3. This means that all the outliers are pulled inward, just to the ends of the distribution.

Third, we calculate the skewness of the distribution of adjusted *Z*-scores by subtracting the mean of the adjusted *Z*-scores from the given adjusted *Z*-score, then dividing the difference by the standard deviation, then cubing the quotient, then adding together all such products (one for each adjusted *Z*-score), and finally dividing the sum by the number of *Z*-scores. In other words, we're using the average cube of the *Z*-score of the adjusted *Z*-score. This is a lot to unpack, but it essentially means that we are trying to tell whether (and if so, to what degree) the distribution is distorted away from a perfect bell curve in either the positive or negative direction. We need to cube the *Z*-scores in order to preserve the distinction between positive and negative skews, because merely multiplying a negative value by itself would result in a positive value and thereby blind us to the direction (see Equation 4):

$$Skew = \left(\frac{1}{n}\right) * \sum_{i=1}^{n} \left(adjusted\ \ Z\text{-}score - \mu(adjusted\ \ Z\text{-}score)\right) / \sigma)^3 \tag{4}$$

Fourth, we calculate the *scaling divisor* (for calculating the interim ratings). The purpose of the scaling divisor is to provide a measure of the spread of the distribution. In the next step, dividing by this spread controls for the differences in the spread between pillars.

We calculate the scaling divisor by multiplying 2 by either the absolute value of the lowest- or highest-adjusted *Z*-score, whichever value is greater. We then round the product up to the next largest integer, as follows:

$$Scaling\ \ divisor = Ceiling\left(2 * \max\left\{\left|\min(adjusted\ Z\text{-}score)\right|, \max\left(adjusted\ \ Z\text{-}score\right)\right\}\right) \tag{5}$$

where the function *ceiling* consists of rounding the value up to the next largest integer.

Fifth, we calculate the *interim rating* for each raw score. This step corrects for the median drift by subtracting out the skewness and median *Z*-score from each individual *Z*-score and adding 0.5 as the new median. As mentioned previously, we also divide the isolated

difference of the individual *Z*-score by the scaling divisor, which makes these interim ratings more comparable across pillars, despite the different degrees of spread.

We calculate the interim rating by subtracting the skewness from the adjusted *Z*-score, then subtracting the median adjusted *Z*-score from the difference, then dividing the resulting difference by the scaling divisor, and finally adding 0.5 to the quotient, as follows:

$$Interim\ rating = 0.5 + \frac{Adjusted\ Z\text{-}score\ -Skew-Median\ adjusted\ Z\text{-}score}{Scaling\ divisor} \quad (6)$$

The sixth and final step of within-pillar calculation is designed to map the ratings in each pillar onto a normal distribution, thus reducing the clustering or crowding of the data points. This is done by dividing the distribution of interim ratings into four quartiles and translating each data point into 1 in a normal distribution according to its placement relative to the limits of its quartile. Intuitively, this process is similar to a painter transferring an image between differently sized and shaped surfaces with the help of a grid.

First, we find the median of the distribution and translate that to 0.5 in the normal distribution. Then, we find the midpoints between the median and bounds (upper and lower) of the original distribution and translate those to 0.75 and 0.25, respectively, in the normal distribution. For each data point, we calculate the ratio of the difference between the normal distribution quartile bounds to the difference between the bounds of the original quartile and multiply that ratio by the sum of the lower bound of the normal quartile and the difference between the interim rating and its original quartile's lower bound, as follows:

$$Rating = 100 * \left(L_1 + (Interim\ rating - L_0) * \left(\frac{U_1 - L_1}{U_0 - L_0} \right) \right) \quad (7)$$

where L_1 and U_1 are the lower and upper bounds of the normal-distribution quartile, respectively, and L_0 and U_0 are the lower and upper bounds of original quartile, respectively. Because the second quartile's upper bound is set at 0.5 and the transfer formula is multiplied by 100, the resulting normal distribution is centered at 50 (with a range of 0–100).

After this sixth step, we are finally ready to publish the three pillar ratings. Those pillar ratings can, in turn, be used to calculate the composite ESG rating, by multiplying each pillar rating by one-third and then adding the three ratings together, as follows:

$$Composite\ ESG\ Rating = \frac{1}{3}(Environmental) + \frac{1}{3}(Social) + \frac{1}{3}(Corporate\ Governance) \quad (8)$$

APPLICATIONS OF ESG RATINGS

Now that we have seen in detail how the TRCRR are constructed, we will briefly explore some ways in which they have been (and may come to be) applied.

We have used the TRCRR as the basis for a family of indices, in order to continuously track how ESG ratings correspond to the movements of domestic and global markets. Since launching in 2013, the Thomson Reuters Corporate Responsibility Indices (TRCRI, developed jointly by Thomson Reuters and S-Network Global Indexes) have served as benchmarks for the performance of the top-ESG-rated companies. As such, they strongly correlate with the ups and downs of the broader market, but often with slight outperformance, perhaps due to an effect of rating-based portfolio restriction on risk-adjusted return, which was found by De and Clayman [2015]. In any case, regardless of whether the indices outperform or underperform the market at a given time, they serve an informative role by tracking the universe of companies from which ESG-conscious investors will likely pick.

The Thomson Reuters Corporate Responsibility Index family comprises 12 region- and/or pillar-specific indices intended as benchmarks for the performance of high-ESG-rated stocks. Exhibit 4 shows the breakdown.

As we can see, the indices are split into three regions of markets: the United States, developed markets excluding the United States, and Europe. They are also split by pillar: In addition to an index of the top stocks by overall ESG rating, each region has indices of the top stocks by each pillar rating.

Each of the indices is created by taking the top half of stocks (by the relevant rating) from an underlying bench-

EXHIBIT 4

Thomson Reuters Corporate Responsibility Index Family Breakdown

Index Name (Thomson Reuters CRI…)	Ticker
U.S. Large Cap ESG	TRESGUS
U.S. Large Cap Environmental	TRENVUS
U.S. Large Cap Governance	TRCGVUS
U.S. Large Cap Social	TRSCUS
Developed Markets (ex-U.S.) ESG	TRESGDX
Developed Markets (ex-U.S.) Environmental	TRENVDX
Developed Markets (ex-U.S.) Governance	TRCGVDX
Developed Markets (ex-U.S.) Social	TRSCDX
Europe ESG	TRESGEU
Europe Environmental	TRENVEU
Europe Governance	TRCGVEU
Europe Social	TRSCEU

mark index published by S-Network Global Indexes: The U.S. indices are drawn from the S-Network U.S. Large Cap 500 Index; the Europe indices, from the S-Network Europe 500 Index; and the Developed Markets indices, from the S-Network Developed Markets (ex-U.S.) 500 Index (SNDMI), which is composed of the largest half (by market capitalization) of the S-Network Developed International 1000 Index (SND1000).

The TRCRI are designed to neutralize style exposures and portfolio growth biases. Because the screens for the indices are based on the TRCRR, they have built-in fair comparison across industries and regions, using a best-of-class approach that avoids over- or underweighting by sector, country, or stock type (e.g., growth vs. value). Sector weights for each TRCRI are neutral with respect to the relevant benchmark index from which the stocks are drawn.

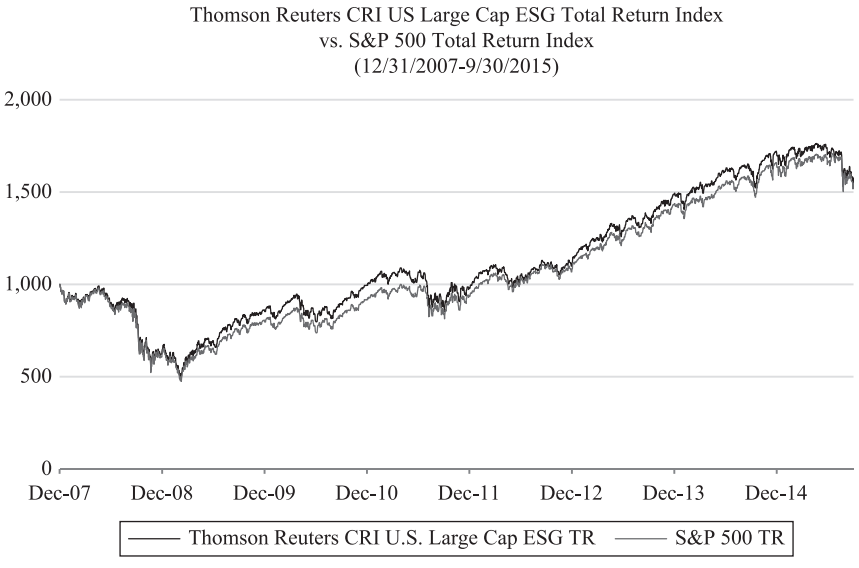
The companies are selected through a combination of variously weighted market capitalizations and ESG ratings, thus striking a balance between high ESG performance and prominence in the market. The results speak for themselves; in Exhibit 5, we show an example of the performance comparison (in this case, the U.S.-based Large Cap Total Return Index compared with the S&P 500 Total Return Index). As of September 30, 2015, the correlation was 0.9897, and the benchmark correlations for the developed markets and European versions were even higher.

We also have recent evidence that the TRCRR can serve to guard portfolios against risk from those companies with the worst practices. De and Clayman [2015] used the TRCRR to study the effects of excluding the lowest-rated companies from a portfolio. They removed the bottom 10% of companies by ESG rating to create a restricted universe. Then, in order to model the approach of an active manager, they assembled 40-stock portfolios using random selection (without replacement) from the restricted and unrestricted universes (100 portfolios from each) and compared the performance of these simulated portfolios. Their results were striking: The restricted portfolios outperformed the return of the unrestricted controls in 71% of the cases simulated.

De and Clayman produced a number of other interesting findings, such as that high ESG ratings strongly correlated with low volatility (a relationship that strengthened as overall market volatility increased); that there was a positive correlation between ESG ratings and stock performance, particularly following the 2008 financial crisis (but none significant at other times); and that the companies with the highest risk-adjusted returns almost always had high ESG ratings (in a methodological reversal of the first analysis we described). There are also

EXHIBIT 5

Performance Comparison



Sources: Bloomberg and S-Network Global Indexes, Inc.

questions to explore within the framework of ESG ratings; in an earlier article, for instance, the same authors found that corporate governance measures were more strongly related to stock returns than are the E or S components of ESG scores (De and Clayman [2010]).[3] This is the sort of richly detailed research that comparable ESG ratings enable, and we hope it continues to flourish.

CONCLUSION

We have seen how the TRCRR are constructed, a few ways they have been used, and some evidence for the effectiveness of these and similar data-driven strategies. Hopefully, this article leaves you with an intuitive grasp of the logic behind the process. We also hope that you now have enhanced confidence in the quality of these ratings, such that you feel confident in using them to study how ESG practices influence stock performance, stability, popular perceptions, and countless other variables.

APPENDIX

EQUATIONS

Metric Score Conversion

For environmental, positive polarity:

$$Score = 0.6 + 0.4 * \left(\frac{x_i - \min\limits_{x \in I} x}{\max\limits_{x \in I} x - \min\limits_{x \in I} x} \right)$$

For environmental, negative polarity:

$$Score = 0.6 + 0.4 * \left(1 - \left(\frac{x_i - \min\limits_{x \in I} x}{\max\limits_{x \in I} x - \min\limits_{x \in I} x} \right) \right)$$

For social or corporate governance, positive polarity:

$$Score = \frac{x_i - \min\limits_{x \in R} x}{\max\limits_{x \in R} x - \min\limits_{x \in R} x}$$

For social or corporate governance, negative polarity:

$$Score = 1 - \left(\frac{x_i - \min\limits_{x \in R} x}{\max\limits_{x \in R} x - \min\limits_{x \in R} x} \right)$$

Weight Calculations

For environmental:

$$reporting < 10\% \rightarrow multiplier = 0$$

$$10\% \leq reporting \leq 30\% \rightarrow multiplier = 0.5$$

$$reporting > 30\% \rightarrow multiplier = 1$$

$$Environmental\ KPI\ Weight = \left(\frac{RLI}{\sum (RLI * Multiplier)} \right) * Factor\ Weight$$

For social or corporate governance:

$$reporting < 0.5\% \rightarrow multiplier = 0$$

$$0.5\% \leq reporting \leq 15\% \rightarrow multiplier = 0.5$$

$$reporting > 15\% \rightarrow multiplier = 1$$

$$Social\ or\ Corporate\ Governance\ KPI\ Weight = \left(\frac{RLI}{\sum (RLI * Multiplier)} \right)$$

Converting Raw Scores to Ratings

Step 1

$$Z\text{-}score = (raw\ score - \mu(raw\ scores)) / \sigma(raw\ scores)$$

Step 2

Right tail:

If $Z\text{-}score > 3$, $adjusted\ Z\text{-}score = \max(Z\text{-}score \leq 3)$

Left tail:

If $Z\text{-}score < -3$, $adjusted\ Z\text{-}score = \min(Z\text{-}score \geq -3)$

Step 3

$$Skew = \left(\frac{1}{n} \right) * \sum_{i=1}^{n} \left(adjusted\ Z\text{-}score - \mu(adjusted\ Z\text{-}score) \right) / \sigma)^3$$

Step 4

$$Scaling\ divisor = Ceiling\left(2 * \max\left\{ \left| \min(adjusted\ Z\text{-}score) \right|, \max\left(adjusted\ Z\text{-}score \right) \right\} \right)$$

where *ceiling* rounds the value up to the next largest integer.

Step 5

$$Interim\ rating = 0.5 + \frac{Adjusted\ Z\text{-}score - Skew - Median\ Adjusted\ Z\text{-}score}{Scaling\ divisor}$$

Step 6

$$Rating = 100 * \left(L_1 + (interim\ rating - L_0) * \left(\frac{U_1 - L_1}{U_0 - L_0} \right) \right)$$

where L_1 and U_1 are the lower and upper bounds of the target rectangle, respectively, and L_0 and U_0 are the lower and upper bounds of the interim rectangle, respectively.

Composite Score Calculation

$$Composite\ ESG\ Rating = \frac{1}{3}(Environmental) + \frac{1}{3}(Social) + \frac{1}{3}(Corporate\ Governance)$$

ENDNOTES

[1]We cannot say exactly how similar in methodology, as rating providers vary in how much of the process they detail in publicly available fact sheets or rule books (and it obviously impractical to publish exhaustive methods, given the qualitative nature of the basic data points). But most ESG rating systems are broadly similar in that they incorporate data on ESG practices.

[2]A consequence of the way the KPI weights are calculated is that all the weights for the KPIs within a peer group (be it industry, region, or universe) sum to one.

[3]Because this study was conducted in 2010, the authors' resources were more limited; they constructed their own ratings using data from KLD STATS (since acquired by MSCI).

REFERENCES

Adler, T., and M. Kritzman. "The Cost of Socially Responsible Investing." *The Journal of Portfolio Management*, Vol. 35, No. 1 (2008), pp. 52-56.

Blank, H. "The ESG Decision Tree." Thomson Reuters, June 20, 2013.

De, I., and M. Clayman "Are All Components of ESG Score Equally Important?" *NYSSA Finance Professionals' Post* (July 2010).

——. "The Benefits of Socially Responsible Investing: An Active Manager's Perspective." *The Journal of Investing*, Vol. 24, No. 4 (2015), pp. 49-72.

Eccles, N.S., and S. Viviers. "The Origins and Meanings of Names Describing Investment Practices That Integrate a Consideration of ESG issues in the Academic Literature." *Journal of Business Ethics*, Vol. 104, No. 3 (2011), pp. 389-402.

Humphrey, J.E., and D.T. Tan. "Does it Really Hurt to be Responsible?" *Journal of Business Ethics*, Vol. 122, No. 3 (July 2013), pp. 375-386.

McQuillen, M.J. ""ESG Portfolios: Changing Beliefs, Perceptions and Goals." *The Future of Investing*, ClearBridge Investments, 2015.

Renneboog, L., J. Ter Horst, and C. Zhang. "The Price of Ethics and Stakeholder Governance: The Performance of Socially Responsible Mutual Funds." *Journal of Corporate Finance*, Vol. 14, No. 3 (June 2008a), pp. 302–322.

——. "Socially Responsible Investments: Institutional Aspects, Performance, and Investor Behavior." *Journal of Banking and Finance*, Vol. 32, No. 9 (2008b), pp. 1723-1742.

Statman, M., and D. Glushkov. "The Wages of Social Responsibility." 2008 Moskowitz Prize Winner; *Financial Analysts Journal*, Vol. 65, No. 4 (July/August 2009) pp. 1-14.

"TRCRR Rule Book." Thomson Reuters, August 14, 2014.

Trinks, P.J., and B. Scholtens. "The Opportunity Cost of Negative Screening in Socially Responsible Investing." *Journal of Business Ethics* (May 2015), pp. 1-16.

Wimmer, M. "ESG Persistence in Socially Responsible Mutual Funds." *Journal of Management and Sustainability*, Vol. 3, No. 1 (2013), pp. 9-15.

To order reprints of this article, please contact Dewey Palmieri at dpalmieri@iijournals.com or 212-224-3675.