JOIM
www.joim.com

JOURNAL OF INVESTMENT MANAGEMENT, Vol. 20, No. 4, (2022), pp. 17–34

ESG INVESTMENT PERFORMANCE EVALUATION: AN INTEGRATED APPROACH

Stephen Horan[a], *Elroy Dimson*[b], *Clive Emery*[c] *and Kenneth Blay*[d]

ESG investment strategies have experienced a massive inflow of capital over the past decade despite investors having few methods to evaluate their performance and communicate their ESG values, objectives, and preferences to investment managers. This paper develops a three-dimensional performance evaluation metric that incorporates return, risk, and ESG outcomes. It is predicated on an investor's willingness to trade off financial gain for non-financial gain and can accommodate any traditional risk-adjusted performance measure. Without such frameworks, investors can neither determine whether outcomes match their expectations nor compare performance across managers and allocate capital accordingly.

1 Introduction

Less than a century ago, investment portfolios were managed across a single dimension—return—and portfolio performance was similarly measured exclusively by return. Markowitz (1952, 1959) introduced risk as a second dimension into portfolio management followed by Sharpe (1964), Lintner (1965), and Mossin (1966), who incorporated it into equilibrium frameworks. As a natural consequence, Sharpe (1966), Jensen (1968), and others developed new portfolio performance measures that integrated both return *and* risk. Other risk-adjusted performance measures followed as our understanding of risk evolved (e.g., Fama and French, 1993, 1996; Carhart, 1996).

Today, more than one-third (36%) of all professionally managed assets, or $35.3 trillion, consider environmental, social, and governance (ESG) criteria according to the 2020 Global Sustainable Investment Review.[1] Signatories to the Glasgow Financial Alliance to Net Zero, a coalition of financial institutions committed to accelerating the decarbonization of the world economy and reaching net-zero emissions by 2050, have grown to 450 firms representing US$130 trillion.[2]

[a]University of North Carolina Wilmington, Cameron School of Business, USA.
[b]University of Cambridge, Judge Business School, UK.
[c]Invesco, Henley on Thames, UK.
[d]Invesco, Boston, MA, USA.

Although such a large proportion of the world's asset managers and assets under management incorporate ESG considerations, the industry has yet to adopt portfolio performance metrics that incorporate all three dimensions of investment performance—risk, return, and responsibility. In fact, it is curious that this massive allocation of capital has occurred without evidence that ESG investing delivers on its promises. A performance metric that incorporates risk, return, and "responsibility" would provide that.

Unfortunately, Gianfrate *et al.* (2021) study the influence of institutional ownership on carbon emissions in 68 countries and find that climate-driven responsible investors are associated with only modest carbon emission improvement. More generally, there is striking dearth of evidence to suggest that ESG investments influence real-world outcomes as investors intend, at least not on a scale commensurate with the asset flows they have enjoyed.[3] Real-world impact from the corporate issuers seemingly most intent on contributing social value, for example, seems lacking. Bebcheuk and Tallarita (2022) examined filings of the over 130 U.S. public companies that joined the much-hailed Business Roundtable (BRT) Statement on the Purpose of a Corporation issued in August 2019 that articulated a more expansive view of corporate stakeholders beyond shareholders. A great majority of the signing firms neither mentioned doing so in their 2020 proxy statements nor mentioned other stakeholders in their corporate governance guidelines. In fact, in response to shareholder proposals regarding the implementation of the BRT Statement during the 2020 or 2021 proxy season, most explicitly stated that their joining the BRT Statement did not require any such changes.

In fact, it isn't clear that investment managers have a sufficiently good understanding of client objectives, which makes evaluating performance against them difficult (Horan *et al.*, 2022).[4] The Financial Conduct Authority (FCA), the financial services and markets regulator in the U.K., notes that ESG/sustainable fund applications lack measurable non-financial objectives.[5] Even for funds that have been approved, it is often difficult to square a fund name or fund objective to its holdings. Gibson-Brandon *et al.* (2021), for example, show that U.S.-domiciled institutional investors who publicly commit to responsible investing have at best the same, or perhaps even lower, ESG scores than institutional investors who do not make a public commitment.

An ESG-adjusted performance evaluation metric that incorporates risk, return, and responsibility would provide evidence that this seismic shift in capital toward ESG investments is worthwhile. It would also allow investors to compare performance over time or across investment managers, thereby holding investment managers accountable for their marketing claims. Frameworks and methods to demonstrate the value of ESG investing in a way that stands up to criticism will do more to benefit people and the planet than glossy marketing brochures.

The next section provides additional motivation and places this work in the context of the existing literature. Section 3 explores investor objectives, which are central to the performance evaluation model that is developed in Section 4. Section 5 explores alternative models and Section 6 concludes.

2 Motivation and Literature Review

Peirce (2018) argues against the concept of stakeholder capitalism on the basis that stakeholders are an ill-defined class of constituents with similarly elastic and sometimes conflicting interests. Shareholders, on the other hand, are a distinct group of individuals whose interests' corporate

managers are intended to serve. Although attractive from a social perspective, this elasticity of identifying stakeholders and their interests makes portfolio evaluation difficult because it impairs clarity, measurability, and accountability.

Performance evaluation of investments with desirable ESG qualities began at least 50 years ago. Bragdon and Marlin (1972), for example, examines the potential tradeoff between financial and non-financial outcomes by relating the profits of 110 virgin paper mills over a five-year period to pollution control activity indices (a precursor to today's ESG ratings) and finds a positive association.

Since then, a plethora of researchers have set out to determine whether (i) socially responsible firms generate higher financial performance[6] or stock returns,[7] (ii) socially responsible funds produce higher risk-adjusted returns,[8] and (iii) socially responsible indices outperform conventional indices.[9] The evidence is inconclusive.

Corporate governance has long been studied as a factor possibly leading to either better financial performance (e.g., corporate outcomes), increased risk-adjusted investment returns, or both.[10] (e.g., Gompers *et al.*, 2003).[11] There are also reasons to believe that social (S) or environmental (E) factors might lead to higher profits, valuations, and/or excess returns. Eichholtz *et al.* (2010), for example, showed that "green buildings" command higher rents and selling prices than otherwise identical buildings.[12] On the other hand, Adler and Kritzman (2008) show that exclusionary socially responsible investing approaches can impose significant costs to investing outcomes, highlighting that there are likely trade-offs between financial and non-financial outcomes investors might desire.

These studies consider performance evaluation in a two-dimensional risk-adjusted return framework. Our contribution is to introduce a three-dimensional performance evaluation framework that incorporates risk, return, and real-world outcomes.

Incorporating the real-world impact of ESG investing into performance evaluation treats real-world outcomes as an investment objective rather than an investment constraint. Over decades, the investment management industry has developed tools like the investment policy statement (IPS) that articulates a client's investment objectives (i.e., return requirements and risk tolerances) and investment constraints. Forty years ago, when responsible investing focused on exclusions, client objectives were expressed in the investment policy statement as a constraint to avoid holdings associated with South Africa, gambling, alcohol, or firearms. In fact, ESG was relegated to the residual bin of investment constraints, called "unique circumstances" (Maginn *et al.*, 2007; Byrne and Smudde, 2019).

In the U.K., responsible investing became more visible in the 1990s.[13] ESG investing is now central to what some investors want to achieve.[14] It has effectively been elevated from a residual investment constraint to a primary investment objective. Unfortunately, we lack mechanisms for clients to express their values and measure performance against them. This paper contributes to the evaluation of ESG investment performance.

3 Client Objectives and ESG Intensity

Evaluating any investment strategy involves comparing end results to original ambitions. Horan *et al.* (2022) introduce a spectrum of interest or commitment an investor has for ESG/sustainable investing ranges from disinterest (characterized by an exclusive focus on traditional financial objectives) to a potentially exclusive focus on non-financial objectives (Figure 1).

Figure 1 Spectrum of ESG/sustainable investing.

Source: Horan *et al.* (2022).

Between those extremes, investors may prefer certain ESG outcomes ceteris paribus but are unwilling to make financial trade-offs to realize them. Their investment manager should therefore consider ESG if, and only if, there are no trade-offs with financial objectives. Others are willing to trade off financial gain for non-financial benefit but perhaps to varying degrees.

We call this willingness, or lack thereof, to make financial–nonfinancial trade-offs, "ESG Intensity". The further to the right an investor places themselves on Figure 1, the greater the intensity, the greater the penchant to make financial tradeoffs, if necessary. Barber *et al.* (2021) call it willingness-to-pay (WTP). Although we measure intensity as the willingness to make financial–nonfinancial tradeoffs, we make no assumption whether ESG factors increase or decrease risk-adjusted returns.

The Horan *et al.* (2022) categories—Traditional, Responsible, Sustainable, and Impact—resemble those identified by a variety of standard setting bodies who have noted the varying degrees of responsible investing (e.g., Sustainable Accounting Standards Board, Financial Conduct Authority). The nomenclature is unimportant, but the continuum uniquely highlights the potential tradeoff between value and values that investors are willing to make. The important implication is that a measure of "intensity" allows investors to communicate to investment managers their willingness to make financial tradeoffs (whether such tradeoffs are necessary in the capital market) and provides a framework on which fund managers and advisors can be evaluated as shown below.

Pederson *et al.* (2021) develop a model of ESG-adjusted efficient frontiers that distinguishes between investor preferences for ESG factors and acknowledges possible financial–nonfinancial tradeoffs. They distinguish between investors who are unaware of ESG factors and those who are aware of them and use them to improve their estimates of risk and expected return. These two classes of investors are like uninformed and informed investors competing to maximize risk-adjusted return. Pederson *et al.* (2021) also define a third group of investors that derives non-financial utility from ESG factors, but they leave the ESG preference function undefined.

The concept of ESG intensity is simple but powerful because it has the following very important uses.

(1) A parameter in client investment objectives, it conveys to the investment manager the importance the client places on ESG considerations.
(2) A parameter in the performance evaluation metric, which would derive directly from client objectives. The greater the intensity factor, λ, relatively more weight is placed non-financial outcomes relative to financial outcomes.
(3) A measure of the appropriateness or aspirational nature of the benchmark. Investors can

assign greater weight (i.e., more intensity) the more confidence they have in the appropriateness of the benchmark as a metric of their intended outcomes. The investor may assign greater weight to aspirational benchmarks because beating them (even in small measure) carries greater impact.

(4) A standardized parameter to compare ESG-adjusted performance across managers.

It also leads us to a definition of ESG, which we discuss now. Statman (2020) argues that the ESG movement lost its way when it conflated financial and non-financial benefits of ESG investing. The transition increased the number of followers, but it degraded the movement from doing good. Selecting investments on ESG factors for financial gain is plausible and interesting. It is, however, no different from fundamental investing.

We differentiate between ESG investing that increases expected risk-adjusted financial returns and ESG investing that is not return enhancing. We make this distinction because traditional investors would anyway select investments and portfolios on ESG factors that are expected to increase risk-adjusted financial return. ESG investing related to "intensity" does not necessarily increase risk-adjusted return but conveys non-financial gains that are also incorporated into investment objectives.

In the next section, we take ESG intensity as a given and focus on the second application above by developing an ESG investment performance evaluation metric.

4 ESG Investment Portfolio Evaluation

Chambers *et al.* (2021) describe successful impact investing as "optimizing the combination of (risk-adjusted) financial return and ESG-related impact" (p. 82). Our approach bears some relation to the optimization framework of Pederson *et al.* (2021), but differs in several respects. First, Pederson *et al.* do not specify an ESG preference function without which an investor has no way of evaluating an investment manager's ESG-adjusted performance. We allow investors to express different intensity preferences for ESG so that advisors can match their financial and non-financial objectives to suitable investment products.

Second, we use the intensity of an investor's ESG preference to evaluate investment performance, a crucial step to meaningfully compare performance of different fund managers across all three dimensions and improve real-world outcomes.

4.1 R^3 Performance evaluation

A hallmark of a strong performance evaluation framework is that it be specified in advance and commonly agreed upon by client and manager to properly set expectations. In other words, investor needs to express their utility function, which in this case includes the traditional economic factors as well as an ESG preference function. In this section, we develop an ESG-adjusted performance evaluation measure that reflects risk, return, and responsibility, called the R^3 ratio.

Agreeing on a properly designed performance metric increases the likelihood that an investor accomplishes their goals. It does, however, put an onus on the investor to articulate those expectations.[15] That is a challenging proposition in today's nebulous ESG environment, but the following performance evaluation framework is designed to make that easier.

The Sharpe ratio is a traditional risk-adjusted performance measure,

$$Sharpe\ ratio = \frac{R_P - R_F}{\sigma_P} \qquad (1)$$

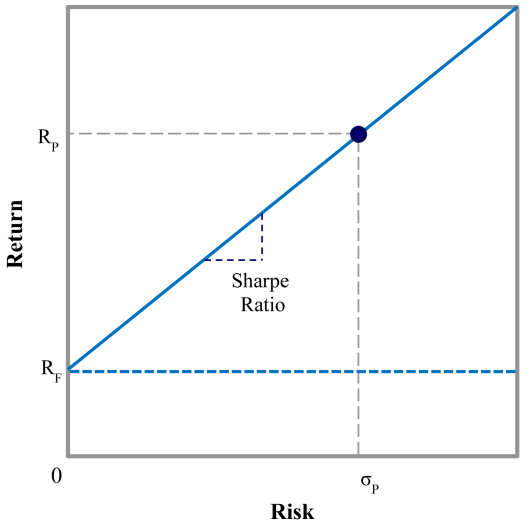

Figure 2 Capital market line.

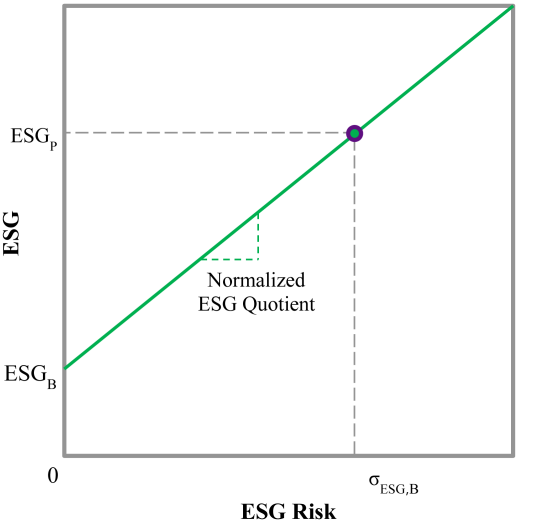

Figure 3 Normalized ESG quotient.

where R_P is the return to the portfolio, R_F is the risk-free rate, and σ_P is the standard deviation of returns for portfolio, p.[16] Figure 2 reminds us that the Sharpe ratio is the slope of the capital market line (CML), which imposes a penalty on the excess return (i.e., $R_P - R_F$) for accepting investment risk above and beyond the risk-free benchmark.

It captures risk and return, but understandably not ESG. In a similar fashion, an investor with ESG intent may desire a portfolio with an ESG profile superior to a particular benchmark and may wish to evaluate the manager's performance on this dimension with a normalized ESG quotient.

In Figure 3, ESG_P is a portfolio-level ESG score according to some pre-determined and mutually agreed upon metric, ESG_B is the benchmark ESG score according to the same metric, and $\sigma_{ESG,B}$ is the standard deviation of ESG scores across the benchmark constituent components.

CFA Institute defines a portfolio-level ESG characteristic as "any measure, or metric, that describes a certain ESG characteristic of the portfolio. A portfolio-level ESG characteristic can be an aggregate measure of the underlying holdings (e.g., asset-weighted carbon intensity) or a measure that is relevant only at the portfolio level (e.g., 85% of assets are invested in green bonds).[17]

We will say more about the choice of benchmark below, but it can represent the market portfolio, a specific industry, or even industry peers. Unlike the standard deviation of portfolio returns (σ_P), the standard deviation of ESG scores across the benchmark ($\sigma_{ESG,B}$) is exogenous.

Its purpose is not to impose a penalty for risk as σ_P functions in the denominator of the Sharpe ratio. Rather, its purpose is to normalize deviations from the mean so that portfolios with large ESG deviations from means with wider distributions are comparable to portfolios with smaller ESG deviations from means emanating from narrower distributions. The aim is to quantify, compare, and illustrate the level of improvement that a fund is delivering relative to its benchmark.

Therefore, portfolios drawing from the same investable universe and having larger positive (negative) deviations from the mean will have a more positive (negative) slope. We call this slope,

the ESG quotient, which is the performance component that captures the third R – responsibility.

$$ESG\ quotient = \frac{ESG_P - ESG_B}{\sigma_{ESG,B}} \quad (2)$$

Assume the investor and portfolio manager agree to evaluate fund performance using a simple Sharpe ratio, which captures two of the three R's – risk and return. If, however, they agree to evaluate performance on the basis of responsibility, as well, the ESG quotient both articulates the investor's objective and adjusts the Sharpe ratio into the R^3 ratio, which reflects all three components—risk, return, and responsibility.

$$R^3 = \frac{R_P - R_F}{\sigma_P} + \lambda \left[\frac{ESG_P - ESG_B}{\sigma_{ESG,B}} \right] \quad (3)$$

where λ is a scaling (or intensity) factor that represents the weight an investor assigns to the ESG quotient based on his or her preferences. Recall, λ performs several functions listed above and which we discuss on more detail below.

In general, the R^3 portfolio evaluation model is consistent with that of Pedersen *et al.* (2021) who accommodate (but do not specify) an ESG preference function that is independent of wealth and variance of wealth. It also aligns with the willingness-to-pay (WTP) tradeoff that Barber *et al.* (2021) document among impact venture capital investors willing to accept 2.5 to 3.7 percentage points lower IRRs in exchange for meeting their dual objectives. An investor's ESG intensity factor would influence the WTP.[18]

4.2 The ESG intensity scaling factor

An investor's ESG intensity is a measure of their willingness to trade off financial for non-financial outcomes. The further to the right on the intensity scale in Figure 1, the higher the scaling factor, the greater the willingness to make tradeoffs, if necessary. A "traditional" investor, for example, on the continuum represented in Figure 1 would assign no weight or intensity to the ESG quotient ($\lambda_{Traditional} = 0$) in which case the investment manager is evaluated purely on risk-adjusted returns.

A "sustainable" investor on that continuum might be willing to make financial tradeoffs for ESG gain, and assign a moderately positive weight with $\lambda_{Sustainable} > 0$ because their ESG intensity and WTP is higher. An "impact" investor might be willing to make even greater financial tradeoffs for non-financial gain and assign more intensity yet, such that $\lambda_{Impact} > \lambda_{Sustainable}$.[19]

The intensity factor, λ, is analogous to an investor's risk-aversion parameter in a mean–variance optimization (MVO) utility function. Although some authors have demonstrated how it can be meaningfully and practically estimated (Wilcox *et al.*, 2006; Horan and Johnson, 2014), advisors rarely use it in practice. As a normalized deviation from the mean, the ESG quotient provides a powerfully intuitive interpretation for the scaling factor that can be used in practice.

This approach makes no assumptions about the relationship, if any, between ESG and risk-adjusted returns. Specifically, it does not presume a higher ESG quotient is associated with a lower Sharpe ratio.[20] The scaling factor, λ, does provide a mechanism however to measure a client's WTP should a tradeoff between financial and non-financial exist or become necessary.

For example, an investor willing to accept a 0.10 lower Sharpe ratio in exchange for a portfolio ESG one standard deviation above the benchmark would express their ESG intensity, λ, as 0.10. An investor willing to accept a 0.20 lower Sharpe ratio in exchange for the same one standard deviation of ESG outperformance is expressing their λ to be 0.20, and so on. Financial markets may not

require that tradeoff, but λ provides guidance to the portfolio about how to evaluate them should they exist.

R^3 is an ad hoc construct. One can conceive of others, and we consider some below. The benefit of this approach, however, is that it provides an intuitive opportunity to express ESG intensity and performance as a possible tradeoff for traditional investment performance expressed in units of the stand-alone traditional performance measure in a way investors can easily understand. Providing the investor with a framework to articulate their investment objectives in this way with some specificity is a major practical advance.

R^3 is so named because it captures three dimensions of investment outcomes—risk, return, and responsibility. We use the Sharpe ratio here as an animating risk-adjusted performance measure that captures the first two R's, but we could substitute any suitable performance metric, of which there are many.

4.3 Benchmark selection

Benchmark selection is critical and can influence the choice of λ. Until now, we have discussed λ as an intensity factor that represents investor preference for ESG results relative to a benchmark (e.g., $\lambda = 0$ for "traditional" fund, low λ for "responsible" fund, high λ for "impact" fund).

The scaling (intensity) factor can also qualify the benchmark used to compare a fund. A client may want to give more "credit" for positive deviations from a more aspirational benchmark and therefore assign greater weight, or intensity. Conversely, an investor might feel less intense about beating a modest benchmark because positive deviations a less ambitious ESG benchmark will produce a larger ESG quotient to which an investor may wish to assign lesser intensity.

Similarly, benchmarks that are either more or less aligned with an investor's ESG goals would warrant different ESG intensity factors. Benchmarks with greater (less) alignment would presumably warrant a larger (smaller) intensity factor. In the extreme, an investor would assign $\lambda = 0$ to ESG benchmarks bearing no relation to their goals. Therefore, benchmark selection is critical and may influence the choice of λ.

A standardized benchmark that reflects the ESG score of a passively managed portfolio, like the market portfolio, has intuitive appeal because it resembles a CAPM framework. It does not enjoy the same theoretical underpinning as CAPM, however. A common alternative benchmarking approach is to compare financial investment performance against that of investment manager peers, such as the average of a peer group or a median manager in a group. So, one could conceivably define the benchmark in the ESG quotient in a similar way. That approach is not best practice, however, because peer groups and median managers do not satisfy many attributes of a valid benchmark because they are neither investable nor specified in advance (Conover *et al.*, 2013). Peer groups and median manager are not investable because the investor cannot replicate them without prior knowledge of the securities and their weightings. Moreover, neither the average, median, nor quartile ranges can be known until the measurement period is over.

One could narrow the scope or raise the bar of a benchmark to a specific industry, such as a low-carbon renewal energy or technology sector, but unless the fund being evaluated had an investment mandate limited to one of those industries it would fail the appropriateness test of a valid benchmark, which requires the benchmark to be consistent with the investment manager's style.

4.4 Advantages

An advantage of the R^3 ratio is that it treats the financial and non-financial components independently. In addition to being consistent with the utility function of Pederson *et al.* (2020), it is also consistent with the notion of distinguishing between ESG factors that increase risk-adjusted return and those that do not. ESG factors producing only non-financial gain are reflected in the second term. Although ESG factors producing financial gain are reflected in both terms, its impact on the second term is determined by the investor's chosen intensity.

Other advantages to this approach are that it is intuitive, offers a framework for the investor to articulate their objectives, and provides clarity for advisors to understand investor intent. These features combine to allow the advisor to lead the client in a product discussion without the quagmire of an ethical discussion.

Another advantage is that it generalizes to any measure of risk-adjusted performance, such as Jensen's alpha, Fama–French alpha, or any agreed-upon ESG metric however nar rowly or broadly defined, such as carbon emissions or ESG rating. Importantly, however, an ESG intensity factor, λ, that properly adjusts the Sharpe ratio will be improper for other measures of risk-adjusted performance, such as Jensen's alpha or the Treynor ratio. So, care must be taken to interpret and choose them accordingly.

4.5 Disadvantages

A potential disadvantage of this approach is it encourages an outsized focus on imperfect metrics. As Cameron (1963) puts it, "Not everything that counts can be counted, and not everything that can be counted counts."[21] Howard-Grenville (2020) emphasizes the importance of focusing ESG metrics on outcomes and impact. She also cautions against fixating on a smaller part of a larger complex system simply because the smaller part can be more easily measured. Doing so risks creating unintended consequences, and she cites CO_2 emissions as a case-in-point.

Muller (2019) also refers to this possibility as the "tyranny of metrics" in his book of the same title, which highlights the pathologies associated with gamesmanship, perverse incentives, and unexpected consequences associated with poorly or overly narrowly designed metrics. Neither Muller, Howard-Grenville, nor Cameron suggests dismissing metrics altogether. They suggest instead placing them in a tapestry of objective measures and subjective judgments that give greater weight to better metrics. We agree and suggest factoring one's confidence in the robustness of the chosen ESG score into the investor's intensity factor, λ.

Another disadvantage is that this framework measures the ESG score at a point in time, which introduces two challenges we discuss in the next section. First, it allows investment managers to window dress their portfolio with high ESG holdings when the score is being measured. Second, it fails to measure investing in tomorrow's transition versus today's virtue. The approach can be modified to address both drawbacks, which we address now.

4.6 R^3 performance evaluation: An example

The R^3 performance evaluation framework can be illustrated with some examples that demonstrate the interpretation of the ESG quotient and how it can be affective by some example ESG strategies.

We construct a 33-stock based portfolio from the MSCI All-Country World Index (ACWI), a large- and mid-capitalization index of 2,966 stocks in 23 developed markets and 25 developing markets. We selected carbon intensity as the ESG

metric because it is quantifiable and the focus of much attention. It is one of many specific metrics an investor could choose, such as methane emissions, gender diversity, or human rights abuses. Alternatively, metrics can be broad—either within one of the E, S, or G categories or across all three.

Stocks are ranked by carbon intensity (Scope 1 and 2 emission per USD sales) within each of the 11 sectors. A 33-stock portfolio is constructed by selecting three stocks from each of the 11 sectors—one stock each from the top decile, the bottom decile, and near the median.[22] The resulting 33-stock portfolio, weighted by market capitalization, is intended to represent a hypothetical "ESG neutral" investment portfolio (see Table 1).

Carbon intensity has a very high variance and is highly skewed. We therefore transform it with the natural log. It remains a figure, however, in which a *lower* score is more desirable from an environmental perspective than a higher score. So, we will adjust our interpretation of the results accordingly. We also treat the 33-stock non-ESG portfolio as the benchmark to isolate the contribution of the ESG strategy. A more typical application would likely be to treat the universe of stocks from which the stocks were selected to be the benchmark.

We design three hypothetical ESG strategies derived from the holdings of the 33-stock non-ESG portfolio, all based on some exclusion method, including:

(1) **Ex-Energy Major**—Divesting the single largest and most carbon intensive energy producer,
(2) **Ex-Energy**—Divesting the entire energy sector, which is a pure ESG asset allocation play, and
(3) **ESG Stock Picker**—Divesting the most carbon intensive holding within each of the eleven sectors.

Table 2 lists the ESG scores for the neutral portfolio and the three ESG strategies. The ex-energy major portfolio improves upon the ESG neutral score by 0.30.[23] Relating that raw differential to the standard deviation of ln(ESG) for the entire universe (2.04) produces an ESG quotient of 0.15.[24] Excluding the entire energy sector improves ESG performance further for a 0.36 raw difference in ln(ESG) and an ESG quotient 0.18.

The more elaborate strategy that excludes the highest emitters from each of the eleven sectors dramatically improves ESG performance by 1.21, which represent a 0.59 standard deviation differential. That these simple strategies all fall well within one standard deviation of the mean illustrates the difficulty of building a portfolio with an ESG metric one standard deviation away from the mean, especially when the standard deviation is so large.

The Sharpe ratios for the ESG neutral portfolio and strategies were computed using 2021 performance data. As a very good year, they are relatively high. We can nonetheless calculate R^3 for each portfolio and gauge the improvement, if any, over the ESG neutral strategy. At a $\lambda = 0.25$, the incremental improvement in R^3 (and the standalone Sharpe ratio) for each of the ESG strategies are 0.18, 0.27, and 0.18, respectively. These differences increase as ESG intensity increases. At a $\lambda = 0.75$, the incremental improvements are 0.25, 0.36, and 0.67, respectively. At a $\lambda = 1.25$, the incremental improvements are 0.33, 0.45, and 0.97, respectively. In other words, R^3 increases with both the investor's ESG intensity, λ, and the breadth and weight of the ESG strategy.

Table 2 illustrates that R^3, the composite performance measure that captures both financial and non-financial dimensions, will vary across ESG Investment strategy and by investor intensity. A simple strategy of excluding a major oil

Table 1 Hypothetical ESG neutral portfolio.

GICS sector	Holding	Weight	Carbon intensity (ESG)	ln (ESG_t)
Communication services	Auto Trader Group PLC	0.42%	0.90	−0.11
	Cable One Inc.	0.49%	19.00	2.94
	Orange Polska SA	0.12%	139.00	4.93
Consumer discretionary	Flutter Entertainment	1.07%	1.70	0.53
	Dollarama Inc.	0.59%	27.50	3.31
	Carnival Corp	0.78%	517.20	6.25
Consumer staples	Lawson Inc.	0.22%	4.10	1.41
	Coca-Cola Company	10.26%	48.30	3.88
	Kimberly Clark de Mexico	0.11%	313.70	5.75
Energy	Schlumberger NV	1.82%	69.30	4.24
	Cenovus Energy Inc.	1.08%	565.40	6.34
	Chevron Corp.	9.86%	613.90	6.42
Financials	PICC Property & Casualty	0.27%	0.10	−2.30
	Bank of Montreal	3.04%	3.60	1.28
	Berkshire Hathaway Inc.	16.34%	274.60	5.62
Health care	Humana Inc.	2.44%	0.90	−0.11
	Merck & Co. Inc.	8.57%	23.60	3.16
	Lonza Group AG	2.71%	270.20	5.60
Industrials	Toyota Tsusho Corp.	0.70%	2.00	0.69
	Caterpillar Inc.	4.74%	36.40	3.59
	Singapore Airlines Ltd.	0.48%	1,453.90	7.28
Information technology	Paypal Holdings Inc.	9.84%	1.20	0.18
	Oracle Corp.	11.24%	15.10	2.71
	ON Semiconductor Corp.	1.20%	556.50	6.32
Materials	Johnson Matthey	0.24%	19.90	2.99
	Grupo Mexico SAB de CV	1.47%	494.30	6.20
	PT Semen Indonesia (Persero)	0.15%	10,995.10	9.31
Real estate	Prologis Inc.	5.05%	1.30	0.26
	Swiss Prime Site AG	0.32%	34.20	3.53
	Digital Realty Trust Inc.	2.16%	767.90	6.64
Utilities	ENN Energy Holdings Ltd.	0.96%	26.10	3.26
	Fortum Oyj	1.15%	821.10	6.71
	Huaneng Power International	0.10%	13,505.40	9.51

Source: MSCI, Invesco.

Table 2 ESG performance evaluation for three stylized ESG strategies—compared to an ESG neutral strategy.

	Ex-energy			
	ESG neutral	Major	Ex-energy	Stock picker
Panel A: Portfolio statistics				
ESG_p	3.66	3.36	3.30	2.45
ESG quotient	—	0.15	0.18	0.59
Sharpe ratio (3-year)	1.08	1.22	1.31	1.31
Panel B: R^3 ESG performance evaluation				
$R^3(\lambda = 0.25)$	1.08	1.26	1.35	1.46
$R^3(\lambda = 0.50)$	1.08	1.29	1.40	1.61
$R^3(\lambda = 0.75)$	1.08	1.33	1.44	1.75
$R^3(\lambda = 1.00)$	1.08	1.37	1.49	1.90
$R^3(\lambda = 1.25)$	1.08	1.41	1.53	2.05

Source: Invesco.

producer adds a measure of non-financial performance above and beyond the Sharpe ratio. This is consistent with Dimson *et al.* (2020) who demonstrate that simple exclusion strategies have little effect on the risk–return profile of an otherwise diversified portfolio. Excluding the entire energy sector has a more significant impact.

4.7 Today's virtue versus tomorrow's transition

The point-in-time ESG measure above suffers from the possibility of window dressing (i.e., loading a portfolio with highly rated ESG firms when the ESG measurement is taken). It will mimic a "green" portfolio at that point in time, but financial performance will be driven by "brown" holdings over the course of time (i.e., the evaluation period). In the extreme, a fund manager could own the worst offenders for 364 days of the year and angels on the measurement date. A way to disincentivize window dressing is to take measurements at more than one point in time.

An important aspect of client objectives is the distinction between investing (divesting) in high (low) ESG firms versus investing in firms expected to have the *greatest increases* in ESG score. It is a fundamental distinction because portfolio managers making active bets on companies that will transition the most or the most quickly (on either an absolute or relative basis) will struggle to increase allocations to, say, high carbon emitting firms expected to cut emissions the most or the most swiftly because ESG_P is a point-in-time measure of "today's virtue" rather than "tomorrow's transition".

Net Zero methodologies are among the numerous requirements being imposed on fund managers to monitor their delivery on non-financial outcomes. An industry wide approach has been to reduce portfolio carbon exposure by 30% from a base year (2019) and commit to a 7% per annum reduction thereafter at the portfolio level, i.e., $ESG_{P,t}/ESG_{P,t-1} - 1 > 7\%$. This seemingly sensible approach creates perverse incentives that will not support real-world change.

Funds could have set their 2019 benchmark year as one with large exposures to the high carbon emitting sectors and simply reduce their exposure to these sectors over time to ensure a mathematical reduction in overall portfolio exposure. It is not a real-world solution because, although it reduces

portfolio exposure to carbon, not owning an asset does not mean the carbon is not being produced. Moreover, it encourages owning companies with the lowest emissions (e.g., technology versus petroleum) more than companies that will reduce emissions the most.

4.8 *Measuring firms that change rather than portfolios that change*

An alternative to this blunt and ineffective portfolio measure is ΔESG_P, which measures year-on-year change in the portfolio ESG score. Measuring ΔESG_P at the portfolio level ($ESG_{P,t}/ESG_{P,t-1} - 1$) does not fully address the concern of window dressing because the manager can still load the portfolio with high ESG holdings at the end of the performance evaluation period. Measuring year-on-year change at the security level (i.e., $\Delta ESG = \frac{w_i \sum_{i=1}^{n} ESG_{i,t}}{ESG_{i,t-1}} - 1$) better mitigates the disadvantage of the ESG score being a single point-in-time measure that does not capture evolution over time.[25] Specifically, it captures how constituent *firms* have changed over the time rather than how the *portfolio* has changed over that time.

This approach can be refined further to address potential of window dressing the portfolio with high ΔESG holdings at the end of an evaluation period by further weighting holdings by the duration of time, d_i, for every holding during the entirety of the evaluation period, such that $\Delta ESG = w_i d_i \sum_{i=1}^{n} ESG_{i,t}/ESG_{i,t-1} - 1$, rather than giving full weight to those holdings at the end of the evaluation period. Whichever method one chooses, it leads to a quotient for the *change* in ESG rather than the *level* of ESG,

$$\Delta ESG\ Quotient = \lambda_\Delta \left[\frac{\Delta ESG_P - \Delta ESG_B}{\sigma_{\Delta ESG,B}} \right] \quad (4)$$

where λ_Δ is the investor's intensity for ESG change. As a result, this performance evaluation framework accommodates either ESG *levels* or ESG *change* without making a case for which objective dominates the other. An investor wishing to invest in transition or change would simply substitute one of the transition metrics, ΔESG, above as the measure of non-financial performance, such that:

$$R^3\ Change = \frac{R_P - R_F}{\sigma_P} + \lambda_\Delta \left[\frac{\Delta ESG_P - \Delta ESG_B}{\sigma_{\Delta ESG,B}} \right] \quad (5)$$

The interpretation of λ_Δ as the willingness to make a tradeoff between financial performance and non-financial performance remains unchanged. The measure of non-financial performance is different, however.

5 Alternative ESG performance metrics

As previously mentioned, R^3 is an ad hoc construct and one can conceive of others. In this section, we introduce some of them and their disadvantages. The R^3 ratio treats traditional financial performance and ESG performance independently, which has several advantages especially since the relationship between is largely unknown. Another approach would allow them to interact. For example, one could define an integrated performance metric, R^{3X}, as:

$$R^{3X}\ Sum = \begin{cases} \frac{R_P - R_F}{\sigma_P} \left[1 + \frac{ESG_P - ESG_B}{\sigma_{ESG,B}} \lambda \right] \\ \quad if\ \frac{R_P - R_F}{\sigma_P} \geq 0 \\ \frac{R_P - R_F}{\sigma_P} \left[1 - \frac{ESG_P - ESG_B}{\sigma_{ESG,B}} \lambda \right] \\ \quad if\ \frac{R_P - R_F}{\sigma_P} < 0 \end{cases} \quad (6)$$

The ESG quotient switches signs if the Sharpe ratio is negative to avoid further penalizing managers with poor financial performance for doing "good" with ESG.

In this multiplicative formulation, the influence of the ESG quotient on the R^{3X} ratio is directly related to the absolute value of the Sharpe ratio, meaning that the greater the absolute value of pre-ESG risk-adjusted performance, the greater the ESG "boost". The poorly performing manager with a positive, but low, Sharpe ratio would receive a smaller ESG "boost" for the same level of ESG performance as a manager with a higher Sharpe ratio. We cannot postulate a convincing rationale for that feature.

In fact, the influence of ESG tends toward zero as the Sharpe ratio tends toward zero. It is not clear why a portfolio manager earning the risk-free return (either with or without risk) would receive no credit for constructing a portfolio that beat its ESG benchmark.

More generally, the poorly performing manager with a non-positive Sharpe ratio has no opportunity to convert her non-positive Sharpe ratio into a positive R^{3X} ratio with positive ESG performance. In other words, no amount of ESG "goodness" can compensate for a non-positive Sharpe ratio. Some investors defined as "responsible" or perhaps even "sustainable" on the continuum in Figure 1 may find this feature desirable. An "impact" investor in unlikely to find it desirable, though.

In this case, the choice of intensity ratio would be fundamentally different. For example, an investor may wish to quantify the R^{3X} "reward" for a portfolio that achieves an ESG score in the upper quartile of ESG scores. According to Chebyshev's theorem, the upper quartile is $\sqrt{2}$, or 1.41, standard deviations away from the mean.[26] An intensity factor of $\lambda = 1$ would therefore inflate the pre-ESG Sharpe ratio (or whatever measure of risk-adjusted performance was chosen) by a factor of $1+1.41 = 2.41$ for landing just inside the upper quartile. This heavy weight might align with a non-financial weighting sought by an "impact" investor.

A "sustainable" investor less interested in making financial and non-financial trade-offs may feel a 2.41 multiple for landing in the upper quartile is overly generous and may wish to apply only a 1.70 multiple in which case they would choose to cut the intensity in half and assign $\lambda = 0.5$.[27] A "responsible" investor might feel that intensity still over weights non-financial results over financial results and may choose $\lambda = 0.2$, which would inflate the financial returns by a factor of 1.28 for landing in the upper quarter of ESG scores. Either way, it is a subjective parameter based on investor preference.

A related construct is to define the performance metric so that the influence of ESG performance is determined by the absolute value of the Sharpe ratio, or:

$$R^{3ABS}Sum = \frac{R_P - R_F}{\sigma_P} + \lambda \left| \frac{R_P - R_F}{\sigma_P} \right| \times \frac{ESG_P - ESG_B}{\sigma_{ESG,B}} \quad (7)$$

This formulation is similar to Equation (6) in that it has the same potential disadvantage that the influence of the ESG quotient is directly related to the absolute value of the Sharpe ratio and tends toward zero as the Sharpe ratio tends toward zero. Unlike Equation (6), however, some amount of ESG "goodness" can compensate for a non-positive Sharpe ratio.

We consider these models inferior to the simplicity of the R^3 Ratio defined in Equation (3) or the R^3 Change ratio defined in Equation (5).

6 Conclusion

ESG investing has experienced a massive influx of capital. Investors apparently want to do good but have few ways to articulate these non-financial objectives to investment managers, hold them accountable for their claims, or compare the holistic performance among managers. Investors may choose to allocate their capital differently if they had methods to determine whether their environmental and/or social ambitions have been realized.

We introduce several investment performance measures based on an investor's willingness to trade off financial gain for some amount of non-financial ESG gain. That tradeoff may not be necessary in capital markets (ESG investing could even create financial gain), but an investor's willingness to make that tradeoff provides guidance to the investment manager to make investment decisions. Alternatively, it can be articulated in an ESG fund's investment mandate to set appropriate investor expectations.

Our model presumes that the investor and investment manager agree on a measure of ESG performance for the ESG quotient that measures performance against a benchmark. Such metrics do not yet exist or at least are not widely used. That observation is an important insight from this work. This work highlights the importance of developing such a measure without which integrated performance evaluation is difficult, if not impossible.

This topic is ripe for further investigation. Work will no doubt continue to decipher which, if any, ESG factors are risk-adjusted return enhancing. Those efforts would continue with or without our contribution here. This paper may inspire future research that focuses on other performance metrics that (1) are more tightly connected to an investor's utility function, (2) are applicable in different investment settings, like non-traditional performance measures used for alternative investments (e.g., money-weighted return, maximum drawdown, Calmar ratio), or (3) are less dependent on a metric of ESG outcomes.

Alternatively, future research might estimate empirically investor "ESG intensity", a critical parameter in the evaluation process. Finally, future research can develop ESG metrics that can be used as portfolio benchmarks and calculated for target portfolios. That advance would enable researchers to measure and compare performance empirically as well as develop and apply performance attribution models that can distinguish luck from skill.

Performance evaluation is a critical element to ensure that investors get what they pay for and can compare performance across managers. This paper is an important step in that direction.

Endnotes

[1] 2020 Global Sustainable Investment Alliance, 2020, Global Sustainable Investment Review.

[2] The Glasgow Financial Alliance for Net Zero, 2021, *The Glasgow Financial Alliance for Net Zero: Our Progress and Plans toward a Net Zero Global Economy*.

[3] Although we are aware of a number of studies showing that carbon emission disclosure mandates decrease reported carbon emissions (e.g., Jovenot and Krueger, 2019; Tamer, 2019; Rauter, 2017) and that other disclosure mandates have been shown to decrease water pollution, increase worker safety, and decrease corruption (Chen *et al.*, 2018; Christensen *et al.*, 2017; Rauter, 2017), there is a dearth of evidence that shifts in investment policy have similar impact.

[4] The Securities and Exchange Commission (SEC) in the U.S. attempts to define ESG investment funds in a 26 February 2021 investor bulletin noting that, "An ESG fund portfolio might include securities selected in each of the three [E, S, and G] categories—or in just one or two of the categories. A fund's portfolio might also include securities that don't fit any of the ESG categories, particularly if it is a fund that considers other investment methodologies consistent with the

fund's investment objectives." Such a description highlights the breadth and lack specificity of the ESG or sustainability monikers.

[5] Financial Conduct Authority letter to authorized fund managers (19 July 2021).

[6] Margolis *et al.* (2009), Friede *et al.* (2015), Busch and Lewandowski (2018), Eccles *et al.* (2018), and Giese and Lee (2019).

[7] Krüger (2015), Halbritter and Dorfleitner (2015), and El Ghoul and Karoui (2017, 2019).

[8] Renneboog *et al.* (2008) and El Ghoul and Karoui (2017, 2019).

[9] Schröder (2007) and Dimson *et al.* (2020).

[10] Becht *et al.* (2003), Denis *et al.* (2003), and Claessens and Yurtoglu (2013) offer surveys of academic studies of the link between corporate governance and firm value.

[11] They develop a corporate governance index based on shareholder rights provisions and find that in the 1990s firms with stronger shareholder rights provisions have higher valuations (i.e., Tobin's q ratios). It is possible, however, that savvy investors understand the opportunity and compete away the profits from investing in well-governed companies. Gompers *et al.* (2003) find, however, that firms with stronger shareholder rights provision led to higher risk-adjusted returns. A combination of higher valuation and higher future excess returns for firms with stronger governance suggests that the value of corporate governance is not (or at least was not) fully priced in by investors. Bebchuk *et al.* (2013) replicate Gompers *et al.* (2003) using more recent data and find that, although good governance firms continued to trade at higher valuations in the 2000s, the link between corporate governance and abnormal returns broke down during that period. This finding suggests that, although well-governed firms may enjoy higher valuations, lower cost of capital and higher profits, investors have learned to properly price this aspect of governance thereby eliminating the profit opportunity in an investment strategy employing this factor. Sloan (1996) develops a different measure of corporate governance based on the aggressiveness of a company's accounting choices as evidenced by its accruals and found that firms with less aggressive accruals produced positive abnormal returns.

[12] Although some of the price premium can be attributed to energy savings, higher rents suggest that the label itself affects perceptions in the marketplace.

[13] Acharya and Dimson (2007) report in their book, *Endowment Asset Management: Investment Strategies in Oxford and Cambridge,* that by 2000 British pension funds were required to record their policy on social and environmental issues, and that British charities were free to adopt an "ethical" investment policy.

[14] Although a minority of investors and assets are committed to an ESG primary investment strategy, many more integrate into their investment program.

[15] An economist might say that the investor needs to express their utility function, which in this case includes the traditional economic factors as well as an ESG preference function.

[16] There are dozens of other performance measures designed to measure different things and adapt to various investment settings. The Sharpe ratio is rarely used to evaluate performance in private equity because illiquidity biases the measure of standard deviation downward. Since-inception internal rate of return (SI-IRR) is a generally accepted alternative in that setting. That said, our framework can accommodate mostly any pre-ESG measure of performance.

[17] CFA Institute (2021b). "ESG Disclosure Standards for Investment Products", Exposure Draft, May 2021.

[18] El Ghoul and Aymen Karoui (2017) demonstrate a similar WTP in a sample of 2,168 U.S. equity funds over the period of 2003 to 2011 in which high-CSR funds display poorer performance, stronger performance persistence, a weaker performance–flow relationship.

[19] In theory, an investor with nefarious ESG intent could assign $\lambda < 0$, but that is unlikely. The more likely scenario is that an investor interested in, say, "sin" investments believing that they are undervalued and offer a superior risk-adjusted return. In that case, their interest in poor ESG scores is purely a consequence of pursuing financial gain—a means to an end rather than an end unto itself—in which case the R^3 performance evaluation would assume $\lambda = 0$.

[20] By contrast, the Sharpe ratio and most other performance measures assume a specific link between risk and return.

[21] Cameron, William Bruce (1963). *Informal Sociology: A Casual Introduction to Sociological Thinking*.

[22] We exercised some discretion in selecting within the top and bottom decile to avoid emerging market and small stock biases. We also ensured that the inclusion of a large petroleum company (in this case Chevron) to illustrate the impact of a single stock exclusion.

[23] Recall, a lower score is associated with lower carbon intensity.

[24] The differential is multiplied by negative 1 to convert a positive difference to being desirable.

[25] The two expressions are equal if portfolio holdings are unchanged over the evaluation period but will be different otherwise.

[26] Chebyshev's theorem states that the maximum proportion of observations that are more than k standard deviations from the mean is $1/k^2$. The minimum proportion of observations that are within k standard deviations from the mean is $1-1/k^2$. When $k=\sqrt{2}$ standard deviations, the maximum proportion of observations that fall outside $\sqrt{2}$ standard deviations and the minimum proportion of observations that fall inside are both 50%. Because the inner two quartiles of a ranking represent 50% of the distribution, the upper quarter breakpoint is $\sqrt{2}$, or ~1.41, standard deviations away from the mean.

[27] $[1+1.41(0.5)] \approx 1.70$.

[28] $[1+1.41(0.2)] \approx 1.28$.

Disclaimer

The views expressed herein are those of the authors and do not necessarily represent the views of Invesco.

References

Acharya, S. and Dimson, E. (2007). *Endowment Asset Management: Investment Strategies in Oxford and Cambridge*, Oxford University Press.

Adler, T. and Kritzman, M. (2008). "The Cost of Socially Responsible Investing," *Journal of Portfolio Management* **35**(Fall), 52–56.

Atta-Darkua, V. and Dimson, E. (2018). "Sector Exclusion" in Øystein Thøgersen *Energy Shares in the Government Pension Fund Global*, Oslo: Norwegian Ministry of Finance.

Barber, B., Morse, A., and Yasuda, A. (2021). "Impact Investing," *Journal of Financial Economics* **139**(1), 162–185.

Bebchuk, L., Cohen, A., and Wang, C. (2013). "Learning and the Disappearing Association between Governance and Returns," *Journal of Financial Economics* **108**(2), 323–348.

Bebchuk, L. A. and Roberto, T. (2022). "Will Corporations Deliver Value to All Stakeholders?," *Vanderbilt Law Review* **75** (forthcoming).

Becht, M., Bolton, P., and Röell, A. (2003). "Corporate Governance and Control," In *Handbook of the Economics of Finance*, Vol. 1, Part A. Edited by Constantinides, George M., Harris, Milton, and Stulz, René M. 1–109. Amsterdam: Elsevier.

Bragdon, J. H. and Marlin, J. A. T. (1972). "Is Pollution Profitable?," *Risk Management* **19**(4), 9–18.

Busch, T. and Lewandowski, S. (2018). "Corporate Carbon and Financial Performance: A Meta-analysis," *Journal of Industrial Ecology* **22**(4), 745–759.

Byrne, A. and Smudde, F. E. (2019). "Basics of Portfolio Planning and Construction," CFA Program Curriculum, Level II, Reading 53.

Cameron, W. B. (1963). *Informal Sociology: A Casual Introduction to Sociological Thinking*, Random House, New York.

Carhart, M. (1996). On the Persistence in Mutual Fund Performance, *Journal of Finance* **52**(1), 57–82.

CFA Institute (2021a). "ESG Disclosure Standards for Investment Products," *Exposure Draft*, May 2021.

CFA Institute (2021b). "ESG Disclosure Standards for Investment Products," November 2021.

Chambers, D. R., Kazemi, H. B., and Black, K. H. (2021). "ESG and Alternative Investments," In *Alternative Investments: An Allocators Approach*, 4th edition, CAIA Association.

Claessens, S. and Yurtoglu, B. B. (2013). "Corporate Governance in Emerging Markets: A Survey," *Emerging Markets Review* **15**, 1–33.

Conover, C. Mitchell, B. D., and Carino, D. R. (2013). "Introduction to Benchmarks", CIPM Curriculum, CFA Institute.

Denis, D. K., McConnell, J. J., Ovtchinnikov, A. V., and Yu. Y. (2003). "S&P 500 Index Additions and Earnings Expectations," *Journal of Finance* **58**(5), 1821–1840.

Dimson, E., Marsh, P., and Staunton, M. (2020). "Divergent ESG Ratings," *The Journal of Portfolio Management* **47**(7), 75–87.

Dimson, E., Marsh, P., and Staunton, M. (2022). "Exclusionary Screening," *The Journal of Impact and ESG Investing* **1**(1), Fall, 66–75.

Eccles, R., Kastrapeli, M., and Potter, S. (2018). "How to Integrate ESG into Investment Decision-Making: Results of a Global Survey of Institutional Investors," *Journal of Applied Corporate Finance* **29**(4), 1–137.

Eichholtz, P., Kok, N., and Quigley, J. M. (2010). "Doing Well by Doing Good? Green Office Buildings," *American Economic Review* **100**(5), 2492–2509.

El Ghoul, S. and Karoui, A. (2017). "Does Corporate Social Responsibility Affect Mutual Fund Performance and Flows?," *Journal of Banking and Finance* **77**, 53–63.

El Ghoul, S. and Karoui, A. (2019). "Fund Performance and Social Responsibility: New Evidence Using Social Active Share and Social Tracking Error," Working paper,

University of Alberta, Edmonton, Alberta, Canada, 2019. SSRN.com/id=3489201

Fama, E. and French, K. (1993). "Common Risk Factors in the Returns of Stocks and Bonds," *Journal of Financial Economics* **33**(1), 3–56.

Fama, E. and French, K. (1996). "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance* **51**(1), 55–84.

Friede, G., Busch, T., and Bassen, A. (2015). "ESG and Financial Performance: Aggregated Evidence from More Than 2,000 Empirical Studies," *Journal of Sustainable Finance & Investment* **5**(4), 210–233.

Gianfrate, G., Kievid, T., and Nunnari, A. (2021). "Institutional Investors and Corporate Carbon Footprint: Global Evidence," EDHEC-Risk Institute Working Paper.

Gibson-Brandon, R., Glossner, S., Krueger, P., Matos, P., and Steffen, T. (2021). "Responsible Institutional Investing around the World," SSRN Working paper. https://papers.ssrn.com/sol3/papers. cfm?abstract_id= 3525530.

Giese, G. and Lee, L. (2019). "Weighing the Evidence: ESG and Equity Returns," MSCI ESG Research, April.

Gompers, P. A. and Metrick, A. (2001). "Institutional Investors and Equity Prices," *Quarterly Journal of Economics* **116**(1), 229–259.

Gompers, P. A., Ishii, J., and Metrick, A. (2003). *The Quarterly Journal of Economics* **118**(1) 107–156.

Halbritter, G. and Dorfleitner, G. (2015). "The Wages of Social Responsibility—Where Are They? A Critical Review of ESG Investing," *The Review of Financial Economics* **26**(3), 25–35.

Horan, S. M. and Johnson, R. R. (2014). "The Influence of a Family Business on Portfolio Management: An Asset-Liability Management Approach," *Journal of Wealth Management* **17**(1), 14–30.

Horan, S., Dimson, E., Emery, C., Blay, K., and Yelton, G. (2022). "The State of ESG Investing: A Portfolio Management Perspective," *The Journal of Impact and ESG Investing*, (forthcoming).

Howard-Grenville, J. (2020). "ESG Impact is Hard to Measure—But it's Not Impossible," *Harvard Business Review*, January 22.

Jensen, M. C. (1968). "The Performance of Mutual Funds in the Period 1945-1964," *Journal of Finance* **23**, 389–416.

Krüger, P. (2015). "Corporate Goodness and Shareholder Wealth," *Journal of Financial Economics* **115**(2), 304–329.

Lintner, J. (1965). "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance* **20**, 587–615.

Maginn, J. L., Tuttle, D. L., Pinto, J. E., and McLeavey, D. W. (2007). *Managing Investment Portfolios: A Dynamic Process*, John Wiley & Sons, Inc.

Margolis, J. D. and Elfenbein, H. A. and Walsh, J. P. (2009). Does it Pay to be Good...And Does it Matter? A Meta-analysis of the Relationship between Corporate Social and Financial Performance, March 1, https://ssrn.com/abstract=1866371.

Markowitz, H. (1952). "Portfolio Selection," *The Journal of Finance* **7**(1), 77–91.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley & Sons.

Muller, J. (2019). *The Tyranny of Metrics*, Princeton University Press; Reprint edition.

Mossin, J. (1966). "Equilibrium in a Capital Asset Market," *Econometrica* **34**, 768–783.

Pedersen, L. H., Fitzgibbons, S., and Pomorski, L. (2021). Responsible Investing: The ESG-Efficient Frontier, *Journal of Financial Economics* (forthcoming).

Peirce, H. (2018). "My Beef with Stakeholders" Remarks at the 17th Annual SEC Conference, Center for Corporate Reporting and Governance.

Renneboog, L., Ter Horst, J., and Zhang, C. (2008). "The Price of Ethics and Stakeholder Governance: The Performance of Socially Responsible Mutual Funds," *Journal of Corporate Finance* **14**(3), 302–322.

Schröder, M. (2007). "Is There a Difference? The Performance Characteristics of SRI Equity Indexes," *Journal of Business Finance & Accounting* **34**(1–2), 1–401.

Sharpe, W. F. (1964). "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance* **19**, 425–442.

Sharpe, W. F. (1966). "Mutual Fund Performance," *Journal of Business* **39**(S1), 119–138.

Sharpe, W. (1964). "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance* **19**, 425–442.

Sloan, R.G. (1996). "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?," *The Accounting Review* **71**(3), 289–315.

Wilcox, J., Horvitz, J., and DiBartelomeo, D. (2006). "Investment Management for Taxable Private Investors," *Research Foundation of CFA Institute*, Charlottesville, VA.

Keywords: ESG; ESG investing; values investing; responsibility; responsible investing; performance evaluation; performance measurement; portfolio management