Abstract

Environmental, Social and Governance (ESG) information demands attention within the asset management industry since it has become widely accepted that making an allowance for ESG criteria within an equity portfolio enhances returns. We test this proposition by incorporating ESG criteria into a worldwide market neutral portfolio using an ‘off-the-shelf’ third party database of individual securities’ ratings. Our results show that incorporating ESG information into a worldwide equity market neutral portfolio yields no additional return, as any benefits from tilting towards a better rated ESG portfolio is already wholly captured by other well-known equity factors. Doing so, however, does not hurt returns. We conclude that ESG should not be considered as a unique equity factor.

Keywords: ESG ratings, Quantitative Asset Management, Equity market neutral, Risk premia

1. Introduction

ESG (Environmental, Social, and Governance) targeted investment strategies have exploded in popularity. The Global Sustainable Investment Alliance (GSIA) estimate a figure of 22.9 trillion dollars being managed worldwide under the auspices of “responsible investment strategies” - a 25% increase since 2014 and representing 26% of all assets under professional management [1]. Investors overwhelmingly believe that sustainable investment will grow even further in significance [2], and, is in keeping with other studies and insights that show asset managers (AM) are being put under increasing pressure to disclose their policy on sustainable investment practices - both directly by clients, and through industry peer pressure [3].

The evolution of ESG within the asset management industry at large, and the quantitative asset management industry in particular, necessitates insight into how ESG information may be captured in a quantitative, systematic fashion, and, what (if any) alpha generating properties it contains after controlling for common, well-known risk factors.

This paper contributes to the existing, and mounting body of literature, by taking the distinctive approach of installing ‘off-the-shelf’ ESG information from a reputable and oft used third-party data provider in a systematically integrated fashion. We refrained from singling out any one pillar of the ESG trifecta, nor did we limit our securities universe by any pre-screening process, or on any pre-determined research question on ESG. Our study, unlike many others, investigates holistically the effect of incorporating all three ESG pillars into a large, worldwide investment pool over a ten year period.

Our main results can be summarized as follows:

1. ESG scores exhibit a large capitalization (large-cap) bias
2. ESG scores might have a developed region bias (because of a country adjustment provision in the data aggregation)
3. ESG as a predictor is negatively correlated to Size, and slightly positively correlated to Low Volatility / Low Beta (LB)
4. An equity market neutral portfolio constructed with ESG ratings as a predictor has a flat world-wide performance (marginally positive in Europe but negative in the US. None of the results are, however, statistically significant)

5. The total performance can be explained by (negative) Small-minus-Big (SMB), (negative) Momentum, and (positive) LB: the performance of the remaining unexplained performance is flat

6. Breaking down into E, S, and G shows that: E and S are compatible with noise. G seems at first to have a positive performance. However, this positive drift is well explained by its correlation with the profitability (Quality) factor.

In the next section we briefly review a selection of academic and industry research papers and share our thoughts on the state of the ESG landscape and the pitfalls of ESG-directed stock selection. We follow by describing the data we used in our analysis in section 3. In section 4 we analyze the predictability of ESG scores and its correlation with other well-documented and understood equity risk factors. In section 5 we focus on the performance of a market neutral world-wide ESG strategy and also look at the three individual E, S, G scores separately. We summarize our conclusions and thoughts in the final section 6.

2. ESG and its Applications

2.1. The Emergence and Drivers of ESG

Interest in ‘socially responsible investing’ (SRI), and by extension ESG, first gained significant traction in the 1980s, as the publication of SRI-themed academic articles increased dramatically (along with interest from main-stream media newspapers and magazine sources) [4]. Practitioners, academics, and the media alike have all caught on to the acronym ESG (along with SRI and “Corporate and Social Responsibility” (CSR)) in designating a field of investment that considers social good in addition to financial return when crafting an investment strategy. The trend in the scale and scope of ESG demand has been sustained, with asset owners (AO) either already allocated to dedicated products that include an ESG-tilted process (one in five), or exploring ways of augmenting their existing exposure (67% of AO “looking to increase their allocation to ESG funds in 2018” [1]).

This unrelenting growth in ESG investment demand stems, according to a report by the Forum for Sustainable and Responsible Investment (USSIF)[1], fundamentally from “client demand”: an assertion that is on par with other studies that found that client demand is the leading driver of interests in ESG, along, in descending order, with risk or reputation management, and, regulatory or legal requirements [5]. A consideration that ESG may “play a key role in broader financial performance” only came in at number four on the list of drivers - highlighting a dispute in the broader ESG debate of the belief that ESG integration, and or making decisions based on ESG factors or metrics, provide investors with increased risk-adjusted returns. Most, nevertheless, subscribe to this belief, and the majority of academic research (see for example [6], [7], [8]) supports the notion that some form of ESG integration has a positive relation to financial market performance. This belief, reinforced by supportive academic research (despite contradictory evidence ([9],[10],[11] amongst others), further bolsters the demand-case for ESG-based investment.

2.2. State of Play and How to Integrate ESG metrics? (If at all...)

Eccles and Viviers [12], having made a study of the dizzying array of opinion and nomenclature suggest defining responsible investment as “investment practices that integrate a consideration of ESG issues with the primary purpose of delivering higher-risk-adjusted financial returns.” Paradoxically, their definition is in and of itself imbued with definitional stumbling blocks, as there is a wide range, and varying degrees

---

1See Deutsche Bank Alternative Investment Survey, February 15, 2018
of ‘integration’ [13].

Perhaps then unsurprisingly a survey by Eccles and Kastrapeli [14] revealed that a mere 21% of institutional investors use “Full ESG Integration”, whereas the vast majority of ESG ‘integration’ mainstream finance amounts to the use of arbitrary screening practices, mostly exclusionary or negative (47% of institutional participants), which involves eliminating worst-in-class performers, and / or involves the deliberate exclusion of companies in certain industries or sectors for ethical, or moral considerations - often called “sin stocks” [15]. While these simple exclusionary practices are common, it may incur a special ‘cost’, i.e. a forfeiture of returns that is referred to as the ‘sin premium’ [16].

A surge in the availability of ESG research and data, from and within academia and industry sources - notably rating agencies - has provided the possibility of more studious research, above and beyond the pedestrian exclusion processes that are still largely the status quo. The availability, and increased sophistication of this data has made an ESG approach to investing a candidate for a quantitative support method in the decision making processes that follow SRI guidelines [13]. It has made an assessment of alpha generation feasible, and has further fanned the visibility of its providers and the interest and profile of sustainable investment.

Our study aimed to clarify whether, since arbitrary exclusion practices seem to yield no Alpha, does “investing with a systematic and explicit inclusion of ESG risks and opportunities in investment analysis” [14] yield better results?

3. Data

The MSCI ESG database, containing monthly ratings for a total of 16,799 worldwide companies from January 2007 until October 2017, was chosen for this study on account of its large coverage; often being used in similar studies; and ranking as one of the two leading ESG data providers [17].

MSCI ESG ratings evaluate each company on a set of “key issues” (key issues being, for example, carbon emissions (in the E pillar), Health and Safety (in the S pillar), and Compensation structure (in the G pillar). The data is collected from various sources (public and from firms via surveys) in order to assess both the exposure of a company to a given issue (exposure score) and to the extent a company has developed robust strategies to manage such levels of exposure (management score). For each issue, the exposure and management scores are combined into a single score ($\in [0,10]$) which is designed to capture whether a company has developed strong enough management practices in order to mitigate the risks coming from its exposure. As a result, a high rating may, for example, indicate either a company with low carbon emissions or a company with high carbon emissions but with a sound strategy to prevent such a level of emissions from becoming a crisis in the future. A low rating will identify companies in which not only the exposure is high, but not enough is being done to mitigate it. With some exceptions, the

---

2Negative or exclusionary screening (also sometimes called ‘Values-Based Exclusions’) is the process of eliminating companies (or in some cases, securities or companies domiciled within or operating predominantly in certain countries) that do not comply with pre-defined standards from the investment universe. Negative screening also includes sector-based exclusions, which consists of eliminating companies from industries such as tobacco, alcohol, and weapons; or companies that are viewed as having destructive social consequences, such as gambling, porn, etc. There are naturally a wide, and widely interpretable swath of companies/sectors/countries that may be excluded based on personal or ethical convictions that may not be universal, nor universally applied. There are however certain international treaties and conventions, such as the Fundamental Conventions of the International Labor Organization (ILO), or laws that prohibit investments in controversial weapons that compel exclusion. CFM for example has a policy of excluding any company involved in the production of controversial weapons.
A preliminary mapping of ESG ratings to stock price time series reveal a large market capitalization (large cap) / low volatility bias that seems to be embedded in the ratings. To highlight this feature we rank stocks in terms of market capitalization (or volatility) and map the rank in a [0,1] interval (0 for small cap or low volatility stocks). We then divide ratings into 10 buckets and plot the average rank as a function of the bucket.

---

3The Global Industry Classification Standard is a standardized classification system for equities. It was jointly developed by MSCI and Standard & Poor’s. See MSCI for further details: https://www.msci.com/gics.

4https://www.msci.com/esg-integration
In Fig. 2 we show the resulting plots. Large cap stocks exhibit, on average, higher ratings - as do low volatility stocks. When regressing (without a constant term) market cap rankings against ESG scores, we find a highly significant positive loading of 0.088 with an $R^2 = 0.66^5$. For volatility, we regress $1 - r$ (where $r$ is the volatility rank) against the ESG score and find a highly significant loading of 0.087 with $R^2 = 0.65$.

![Figure 2: The ESG (overall score) as well as its three components are not independent from market cap and volatility.](image)

We furthermore explore the introduction of a latent bias through the inclusion of a geographical score in the computation of exposure / management scores for individual companies. We did this by measuring the average company ESG score for each given year within each given country. We then recover the GDP per capita$^6$ and finally plot the average ESG score per quantile of GDP in Fig. 3. Since it is reasonable to expect that large cap companies enjoy priority analysis in countries with a lower total coverage, we additionally show results for size-adjusted ESG scores.$^7$

In Fig. 3 we show results for this analysis which highlights this country-bias, namely one ESG point lower for emerging markets with respect to developed markets on average. This is also confirmed by a linear regression of ESG scores against GDP per capita (log), which for both the unadjusted and size-adjusted scores, gives a highly significant loading of around 1 with an $R^2 \sim 0.8$ (0.85 for the size adjusted score). Note that individual company scores also take into account a geographical exposure / management score (at a country or even regional level) which could in principle penalize companies in less ESG compliant countries - whether this is a desirable feature or not goes beyond the scope of this paper.

![Figure 3: The ESG score exhibits a country bias across the combined ESG score, as well as on the individual pillars. Countries with higher GDP per capita typically have companies with higher average companies scores - most notable in the G-pillar. Lower regulatory burdens commonly found and associated in developing countries might explain this observation.](image)

---

$^5 R^2 \to 0$ is insignificant, $R^2 \to 1$ is significant

$^6$GDP per capita (constant 2010 US$) data is obtained from the World Bank. See https://data.worldbank.org/

$^7$Size-adjusted scores are the residuals from a linear regression of scores against the market cap rank (we linearly re-map scores from [0,10] to have comparable ranges.)
4.2. The ESG Factor Explained

The primary goal of our study was to evaluate whether ESG can be considered a unique factor in an equity market neutral (EMN) portfolio. In other words, we want to determine if ESG scores can be measured as a source of excess returns (alpha) above what is already captured by other well-known investment factors (such as size, value, etc.). The methodology of our study can be summarized as follows:

1. We generated an ESG (or E, S and G) predictor \( P \) by ranking stocks according to their MSCI scores. We performed a linear transformation such that \( P \in [-1, 1] \) varies linearly between the lowest ESG score \( P_i = -1 \), and the highest score \( P_i = 1 \).

2. To obtain a pool of additional predictors, we repeated this process for a medley of other well-known strategies, in particular: low volatility (LV), low beta (LB), book value (VALB), profitability (PROF), accruals (ACCR), momentum (MOM), size (SMBC) and turnover (SMB).

3. We constructed a market neutral portfolio \( \pi \) from each predictor using our own proprietary risk model which relies on an adapted mean variance optimization method\[18\]. The results are verified with other familiar portfolio constructions, i.e. dollar neutral (where simply \( \pi = P \)) and beta neutral (as per Frazzini and Pedersen \[19\]).

4. From the market neutral portfolios we construct a P&L time series for each predictor which were used as the input for linear regressions. Unlike other studies, we do not define our universe of stocks based on ESG coverage, but rather identify pools of liquid world-wide stocks: the most liquid 1000 stocks in each of the US and Europe; the most liquid 200 stocks in both Australia and Canada, and the most liquid 500 stocks in Japan. Aggregated worldwide results were obtained by summing different P&Ls with equal weights.

5. Before regressing ESG P&Ls against other factors we partially dealt with the problem of collinearity between regressors with the following recipe: we find clusters of correlated predictors and remove the predictor which is best explained by all the others within each cluster. We identified two clusters of highly correlated predictors: LV-LB and SMB-SMBC. We removed two strategies in each pool (one for each cluster) and were left with a set of 6 reasonably independent predictors (i.e. when regressing one against another, the \( R^2 \) is always below 10\% and residuals have a significant Sharpe Ratio). This “cleaning” procedure does not change the substance of the regression and makes the loadings on different factors more meaningful.

6. We concluded by using our battery of 6 regressors to explain the ESG (or E, S and G) P&L.

5. Results

We plot the worldwide combined ESG, and, separated E, S and G P&Ls in Fig. 4. No significant performance is observed, neither as a combined predictor (Sharpe ratio of 0.17), nor individually. The G-pillar, or governance factor, shows some promise and we plot geographically broken down P&Ls in Fig. 4. We also noted that the correlation across different geographical zones is very low for all the ESG predictors.

We perform a linear regression using our six independent factors and show residuals in Fig. 5. The loadings on different strategies together with their relative p-values\(^8\) are shown in Figure. 6. When running the same analysis separately in each pool, we obtain very similar results and we can only explain a small part of the ESG based P&L’s (\( R^2 \) is low) and what is left has a Sharpe Ratio close to 0. The ESG P&L is mostly explained by a negative exposure to the size factor (confirming results in the previous section) with smaller significant loadings on momentum (negative) and LB (positive). When analyzing the E, S

---

\(^8\)Within a statistical hypothesis test, the p-values determine the level of marginal significance attached to the probability of the occurrence of a given event. The smaller the p-value, the higher the significance (or likelihood that the hypothesis explains the observation), with a pre-defined threshold level of significance - often determined as lower than 0.05 (5\% level of significance) being sufficient to conclude that there exist strong evidence against the null hypothesis.
Figure 4: The market neutral world-wide implementation of a ESG strategy, as well as and including the E, S, and G strategies as separate P&Ls (left). Based on a purely statistical criterion, these P&L are compatible with a zero Sharpe ratio. The only exception might be a Governance-tilted strategy (right).

Figure 5: The ESG strategy explained by classical investment styles. The residual is shown to be non-significant.

Figure 6: Loadings and associated p-values obtained from worldwide linear regressions: ESG (left) and Governance G (right).

and G scores individually, one finds similar results for E and S while the G-pillar seems to offer some signal: the size exposure is smaller and one sees a predominant LB and profitability component (however, also in this case, the residuals have very low Sharpe Ratios).
6. Conclusion and Recommendations

6.1. ESG is not an Equity Factor

We tested the common claim that a portfolio constructed by relying on ESG ratings, that is to say tilted towards securities with a higher ESG ranking, exhibit higher risk-adjusted returns. Academic and industry research strongly suggest, notwithstanding a selection of non-negative and contradictory results, that incorporating ESG metrics may increase risk-adjusted returns. Our results, however, would indicate that any benefit from incorporating ESG credentials into a portfolio is already captured by other well-defined and known equity factors. An ESG-tilted process does not deliver higher risk-adjusted returns, since, once we remove the market cap and volatility bias, ESG as an equity factor has returns compatible with noise (t-statistics of less than 1). Investigating the individual E, S, and G pillars, we found that G is less exposed to Size, with a better P&L (Sharpe ratio = 0.46, t-stat = 1.52). The positive drift is, however, explained by the G pillar’s positive correlation to profitability, i.e. the Quality factor. ESG cannot, as such, be considered as a new, unique equity factor.

Our study contributes to an already immense, and growing body of literature by showing that the systematic integration of ranked, ready-to-use, and unenriched ESG information yields no additional benefit. However, importantly, neither does it seem to affect returns negatively. Managers that consider ‘integrating’ ESG information, relying on third party ESG data providers, could as such feasibly construct such portfolios without sacrificing returns.

Although ESG has garnered more wide-spread acceptance, it would seem that most AMs do not (yet) fully, nor systematically integrate ESG into a decision-making process [20] [21]. While many managers might admit to employing or calling on ESG-metrics to inform their investment decision making, and, to considering full ESG integration as the strategy that is most likely to have an impact on performance as Amel-Zadeh and Serafeim [15] have shown, the practice may not have infiltrated mainstream asset management to the extent that many suggest. Cappucci [22] offers a number of explanations for this state-of-play, chief amongst which is a probable “misalignment of ESG’s long-term benefits and firms’ short-term performance incentives”. He continues to list other concerns and barriers to integration such as (i) data quality; (ii) contradicting measuring standards; (iii) under performance; and (iv) cost. Our findings, notably on the topic of data quality (including concerns of data collection methods, subjectivity, etc.) correspond with Cappucci, and others ([17], [23], [15]). A primary, and widely recognized problem of ESG data (irrespective of the provider) is the relatively short historical coverage. This makes the search for short-term statistically significant patterns far more challenging and seriously prevents an assessment of the long-term profitability of ESG compliant companies (which in principle should be the strength of ESG scores).

We do not contend that ESG-based, or ESG-assisted investment strategies are a fools errand, for, we recognize that many of the current obstacles to ESG investing (primarily that of data legitimacy) may be mitigated in the future. Some have argued that the integration of extra-financial information like ESG will only become mainstream when “there is a shared belief among investors that ESG information is relevant” [24]. The lack of such a shared belief, may be borne out of the mistrust of ESG data, amongst others, but we are committed to remaining cognizant and engaged in the debate on ESG as a predictor for increased long-term returns.

6.2. Final Remarks

Our research has highlighted concerns over the feasibility of the systematic integration of ESG information into a worldwide equity portfolio. For those considering any form of attribution (whether it be a simple exclusionary protocol, or a more sophisticated ratings-based integration process) will necessitate a decision on the data provider - that is to say for those who are not engaged in primary ESG research
(nor want, or need to enrich their own research). Such a judgment would benefit from the formation of a common framework for data providers in order to assess data quality, by satisfying certain minimum criteria (such as minimum coverage thresholds - of global markets as well as sectors; accounting for country biases and regulatory discrepancies; etc.).

We recognize that there are concrete efforts underway to establish universal ESG standards, notably from the Sustainable Accounting Standards Board (SASB) who are developing disclosure standards material to specific industries, organizing them in its Materiality Map. In the absence of legislation or as yet to be adopted industry materiality standards, data providers should be held to certain transparency standards (detailing for example the methodology of collection and compilation), so that end-users may be able to account for the rankings of different data providers and its likely accuracy. This should enable managers to weed out unproven and unreliable data providers, and make sense of the dizzying array of existing and new data providers (there are, by some estimates, 150 providers of ESG research, ratings, rankings and indices [25]).

Asset managers should moreover be able to compare the aggregated ESG ranking of their portfolio against a benchmark employing similar ratings (or a commonly acknowledged industry benchmark). It is conceivable that a portfolio constructed using one data provider will not have similar ESG credentials when compared to a benchmark that was constructed using the ESG rating of a rival data provider. With ESG investment primed to demand even greater importance, these measures should alleviate some of the skepticism and facilitate greater (and speedier) adoption of ESG-driven investment.

7. Endnotes

A special thanks to Jean-Philippe Bouchaud, Laurent Laloux, Philippe Jordan, Guillaume Simon and Matthieu Cristelli for sharing their insightful comments, engaging in many thoughtful debates, and for carefully reading the manuscript.

8. References


---

9See for more details: https://www.sasb.org/materiality/sasb-materiality-map/


