ESG Investment Insights

Benefits and Challenges in Extracting Signals from ESG Data

February 2020

Bloomberg Professional Services

Quantitative Strategies

Bloomberg

ESG Investment Insights

Extracting Signals from ESG Data

AJ Lindeman, PhD
+1 212 617 7491
alindeman@bloomberg.net

Unquestionably investors and asset managers with a mandate to adopt Sustainable Investing principles are moving beyond the use of ESG scores and looking to incorporate direct measures of ESG performance into their portfolio construction processes. This approach is known as *ESG Integration*, and extends the more traditional process of screening based on ESG scores and business involvement.

ESG Integration is meant to rely on specific, granular data in addition to composite scores. However, such data are currently not standardized and often poorly disclosed, especially when considering a global investable universe. In our mind there are several key questions at this juncture, including

- What does the data available today allow us to do? What are the limitations?
- Are best practices being developed?
- Can ESG Integration give an 'edge'?
- Is quantifying ESG signal more about capturing risk or opportunity? How should this affect signal construction?
- Are there dependencies between signal construction and portfolio construction? For example, at what point does one adjust for biases and tilts?

Definitive answers to such questions cannot be determined by a single paper, but our hope is that the following guidance developed herein is found to be useful:

- Established statistical techniques can and should be used to extract signal from data that is sparse and reported with lags.
- Best practices can be borrowed from systematic strategy research as well as from other disciplines, including actuarial science and econometrics.
- Construction of multiple ESG signals can facilitate the creation of strategies that better align with sustainability goals. Such signals need to be based on both quantitative analysis and a value system, such as one embodied in various ESG frameworks. It is not yet possible to defend all decisions in terms of backtesting.
- While it is believed that sustainability principles can be used to identify opportunities, there is not a long historical record of quantitative outcomes to analyze. However, our results do add to existing research suggesting that ESG signals are materially correlated with lower return volatility.
- There are reporting and nominal performance biases that are correlated with company size and varies among industries. Without compensating for these, it is possible to introduce unintended tilts into one's portfolio.

We provide evidence for these conclusions by taking a constructive approach, starting with low-level data and "building up" – avoiding the use of ESG scores entirely. The basis for the results in this paper is a case study involving 280 US companies in resource-intensive industries over the 5 year period from 2013-2017. The data set is one recently enhanced by Bloomberg LP, and is available to clients via the terminal and as a data feed.

Previous papers in this series largely focused on index construction using third-party scoring data. As introduced above, the goal here is to examine ESG data that is new to the index and portfolio construction process.

ESG Frameworks

Frameworks form the basis for identifying ESG signals. They not only inform investors as to what are the most relevant and financially material sustainability issues, but they also indicate to corporations what to disclose. These fields correspond to various sustainability goals and generally reflect inputs and outcomes that are measurable. Moreover, fields are grouped into a hierarchy that serves as an organizing roadmap for signal detection and aggregation. Groups lower in the hierarchy may serve to amplify univariate signal and cancel noise, while groups higher up serve a purpose of reflecting broader ESG themes consisting of multi-factor signals.

ESG risks and opportunities differ by industry, so the reality is that there is no single framework or list of data fields that apply across the board. Commonly used frameworks are listed in Table 1.

Organization	Areas of Focus
GRI Global Reporting Initiative	Offers standards for sustainability reporting on economic, environment and social impacts for a wide range of stakeholders.
SASB Sustainability Accounting Standards Board	Organizes sustainability topics spanning Environment, Social Capital, Human Capital, Business Model and Governance into industry-specific groups with an emphasis on financial materiality.
IIRC International Integrated Reporting Committee	Principles-based approach encouraging both qualitative and quantitative disclosure into financial statements. Standardization is secondary to flexibility and context.
TCFD Taskforce on Climate-Related Financial Disclosures	Develops recommendations for voluntary climate-related financial disclosures in conjunction with the Financial Stability Board with emphasis on scenario reporting.
CDP Carbon Disclosure Project (formerly)	Repository of self-reported data by corporates and cities, focusing on Climate Change, Water and Forests
EU Taxonomy	Developing climate-related disclosure benchmarks and regulations, initially focusing on carbon, greenhouse gasses and alignment with the Paris Agreement.

Table 1. Various frameworks for reporting and organizing ESG data.

Additionally many investment managers and asset owners develop their own frameworks based on their house views of opportunities and risks among ESG factors.

Figure 1 shows a sample ESG framework to give a sense as to the breadth of issues being considered. It is a truncated and simplified collection of factors for demonstrating the main ideas in this paper, covering some representative issues within each of the three ESG pillars. Fully-fledged frameworks can contain nearly 200 individual metrics, with materiality varying by sector and industry.

Additional information concerning frameworks and ESG integration can be found in the references cited in the bibliography.

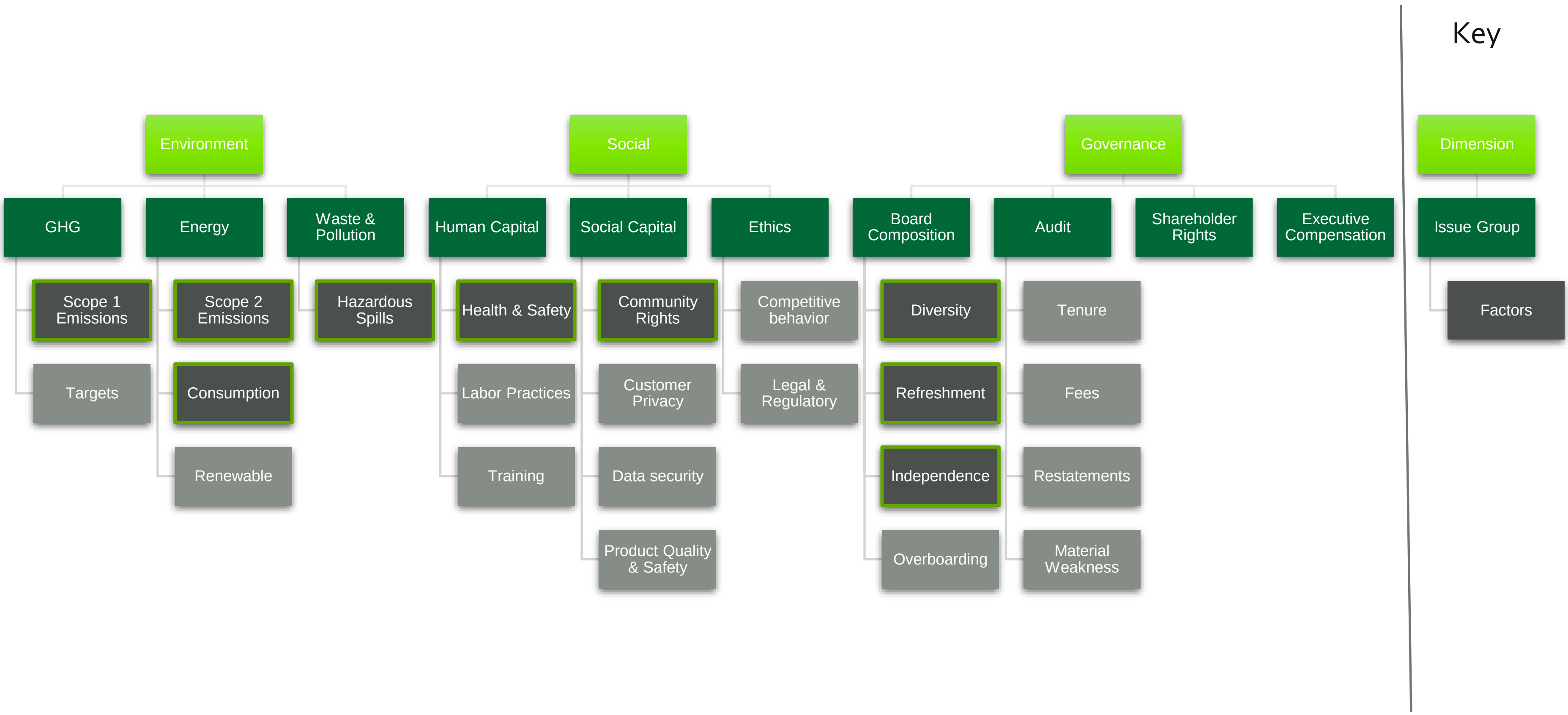

Figure 1: Sample ESG framework. Factors analyzed and incorporated into this study are highlighted with green borders.

Case Study Dataset

Our universe is a set of public US companies in energy-intensive industries. This is so that we have a relatively homogenous set of ESG issues and expected behaviors, which simplifies this expository study.

We observe that each of the 3 sectors has approximately the same number of companies. However, while Energy and Industrials have comparable market capitalizations, Materials has roughly half.

Sector	Industry	Companies	Market Cap
Energy		108	1426
	Oil, Gas & Coal	108	1426
Industrials		94	1678
	Aerospace & Defense	20	698
	Electrical Equipment	26	500
	Industrial Services	8	49
	Machinery	30	343
	Transportation Equipment	10	88
Materials		78	696
	Chemicals	34	412
	Construction Materials	9	59
	Containers & Packaging	14	106
	Iron & Steel	10	43
	Metals & Mining	11	66

Table 2. Summary statistics for companies used in the case study. Sector and Industry classifications are BICS. Market capitalizations are in USD billions as of October 2019. Source: Bloomberg.

To avoid including companies before they became public we required each company's incorporation date to be prior to the end of the reporting period and also applied a heuristic of having 3 or more board members. This results in small variations in the number of companies in the different years within the study.

The time period for the case study is reporting years 2013 through 2017, inclusive. There are significant lags in reporting: as of mid-2019, the majority of 2018 has not yet been disseminated and processed. Furthermore, we emphasize that we work with annual time series. Some fields, particularly in Governance, can change more frequently because they are reported through regulatory filings. However, many fields are only made available on an annual basis through corporate sustainability reports.

Table 3 shows the number of available fields per industry for the 2017 reporting year. Clearly disclosure are higher for Governance than Environmental and Social dimensions: average industry disclosure rates are 99% for Governance, 35% for Social and 23% for Environmental fields. Furthermore, despite our hopes for homogeneity, it is evident that the Hazardous Material Spill fields do not apply to all: various industries, including Industrial Services and Iron & Steel do not have any disclosure. Removing these fields from consideration, the average Environmental disclosure rate is 30%.

		Oil & Gas	Industrials					Materials				
		Oil, Gas & Coal	Aerospace & Defense	Electrical Equipment	Industrial Services	Machinery	Transportation Equipment	Chemicals	Construction Materials	Containers & Packaging	Iron & Steel	Metals & Mining
	GHG Scope 2	21	6	10	1	5	2	20	1	9	0	5
E	Total Energy Consumption	13	6	7	1	5	3	22	1	8	2	3
	GHG Scope 1	27	7	10	1	6	2	20	1	9	1	5
	Amount of HazMat Spills	26	0	0	0	1	0	0	1	0	0	1
	Number of HazMat Spills	24	0	0	0	1	0	7	1	0	0	4
	Indigenous Rights Policy	100	1	1	0	0	0	0	0	1	9	11
	Fatalities - Contractors	23	2	1	2	2	0	12	1	2	1	5
	Fatalities - Employees	24	2	2	2	3	2	14	1	3	2	5
S	TRIR - Contractors	21	0	0	0	0	1	7	1	0	0	1
	LTIR - Contractors	13	0	0	0	0	0	5	1	0	0	1
	TRIR	49	7	7	2	7	4	26	3	8	5	4
	LTIR	23	5	7	2	6	2	16	4	6	1	2
	Human Rights Policy	105	20	24	8	29	10	34	8	12	10	11
	Board Size	105	20	24	8	29	10	34	8	12	10	11
	Chairman Tenure	104	20	24	8	29	10	34	8	12	10	10
	Chairman Age	105	20	24	8	29	10	34	8	12	10	10
	# Board Members Serving > 10Y	103	20	24	8	29	10	34	8	12	10	11
G	# Board Members Serving > 5Y	103	20	24	8	29	10	34	8	12	10	11
	% of Board that are Women	103	20	24	8	29	10	34	8	12	10	11
	Board Average Age	101	18	23	8	29	10	34	8	12	10	11
	Oldest Director Age	101	18	23	8	29	10	34	8	12	10	11
	Youngest Director Age	101	18	23	8	29	10	34	8	12	10	11
	% Independent Directors	104	20	24	8	29	10	34	8	12	10	11

Table 3: Summary statistics for data field availability per field and industry in 2017. Numbers indicate the number of tickers for which a data point is available in the Bloomberg data set.

Computational Framework

Visually, if we turn the ESG framework tree on its side, we get a picture that looks similar to a feed-forward neural network – the difference being that nodes connect based on the ESG framework rather than all nodes connecting to all others. This view offers motivation and rough guidelines for determining the role of each layer.

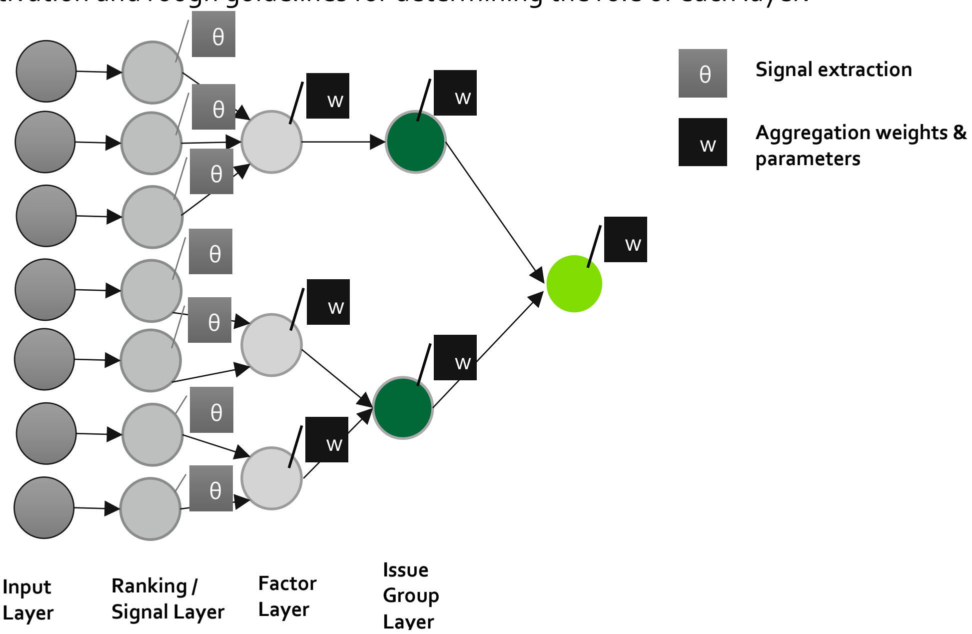

Figure 2: The connection between the ESG framework and computational framework. The conversion to a zero to one signal corresponds to the machine learning concept of applying activation functions to the raw input data. Of course, activation functions usually require parameters, which in the first (univariate) layer are denoted by θ. Common parameters for subsequent layers include weights and are labeled w.

Role and guidelines for univariate signals

We need to translate raw data, which may be in various units, to a standard reference frame. The leftmost nodes in Figure 2 represent raw data. In the first layer, we transform and normalize univariate data – that is, one input and one output. We use a zero to one scale, where naturally the value of 1 is interpreted as "good", that is, indicating that some aspect of an ESG goal is being fully met.

There are two basic approaches to this first signal layer. In a few cases, and depending one's purpose for measuring ESG signal, it can make sense to define the signal curve in terms of external performance guidelines. Gender Diversity and Fatalities are some of the most natural signals for such treatment: sustainability investors do not need to "learn" target levels for gender diversity or fatality rates are from the data. Similarly, it is straightforward to assign signal values of 0 or 1 to fields that indicate whether or not a company has a Human Rights policy without quantitative analysis. That said, for ESG integration specifically, it is open to debate as to whether or not relative performance versus peers is also relevant. In particular, if the investment objectives include tracking error constraints, it becomes necessary to identify the better performers within a peer group. Note that one can convert signals tied to external standards to relative performance simply by ranking within peer groups, whereas the reverse is not possible.

However, when there are no widely-accepted external standards, it is natural to base signals on represent percentiles. This requires a determination of peer groups up front. Groups could be formed by industry, company size, geographical region and several other firm characteristics. It is also possible to apply clustering algorithms, which allows for the data to 'reveal' clustered behavior.

Note there are advantages to using large peer groups, including setting the peer group to be the entire investible universe. Larger peer groups facilitate broader comparisons – something that is of interest to many sustainability investors. But from a traditional mean-variance portfolio construction vantage point, it is actually *essential* to have un-normalized indicators of future risk and performance: one can accept higher levels of (ESG) risk if the return can be expected to compensate for that risk.

Once one settles on desired peer groups, there are two basic choices for estimating percentiles. First one might decide on a non-parametric approach and use the empirical quartiles (or some other quantile). One benefit of doing so is parameter reduction: computing these quantiles only relies on sorting (and the choice of quantile). One downside of using quantiles is that they depend on all the data and are noisy when there are few observations (small peer groups or poor disclosure). Moreover, one cannot compute quantiles until *all* the data for a reporting year is available, which can substantially delay signal determination. Additionally, quantiles de-mean the signals – the median signal value will be 0.5 time after time, making trends somewhat obscured. Moreover, there is risk that clusters of outliers may corrupt quantile's bucket boundaries.

As a remedy to these practical problems, one can choose a parametric distribution that closely approximates the empirical distribution. Such parameters can often be estimated in a robust way, as to not be too sensitive to outliers. These parameters might also be fixed or slowly adjusted over time so that signal trends can be observed. A second benefit of a parametric approach is that it facilitates signal definition when there are very few observable points. One may "borrow" parameters from similar peer groups, or maybe even jointly estimate the parameters of several groups taking into account prior beliefs of each group's mean value.

We emphasize that the interpretation of parametrically fitted signals remains the same as if bucketed quantiles were used. The parametric approach is intended to be a more robust implementation that deals with material issues that arise in practice, and would give the same results as buckets if data were less sparse and noisy.

One implication of either form of percentile estimation is that the high values may not correspond to a sustainability goal. For example, if an entire industry has high pollution levels, does it make sense for the best relative performer to receive a signal near 1? One remedy is to construct peer group signals as well as signals for companies within each peer group.

The role and guidelines for aggregation layers

After the rightmost layer in Figure 2, nodes start to combine. Although these too can be thought of as activation functions, we call this *signal aggregation* because they take multiple inputs. The parameters for these functions are labeled w for weights than can reflect the relative importance among input signals.

While it is true that weights are an aspect of aggregation, there is no reason to limit ourselves to linear functions like weighted averages. If, for example, there is more than one way to meet a sustainability goal, taking the maximum value might be appropriate. In other cases, risks might be best reflected by taking the minimum and passing the worst signal forward to the next level. Thus there are other choices to consider besides weights.

We will discuss these alternatives in greater detail after extracting univariate signals in our case study.

What about machine learning?

While the machine-learning-like signal construction and aggregation flow diagram is meant to provide some comfort that we are taking a structured approach, we do not believe it is feasible to apply conventional machine learning techniques at this point in time. This is because the data is insufficient in terms of historical length and breadth – we do not have "plenty of data". Moreover, historical ESG data may be especially non-representative of future results: climate change in particular is subject to new regulations and changes in aggregate consumer behaviour. Also relevant is the fact that there is no clear objective function for machine learning algorithms to optimize. In fact, financial performance and sustainability goals are measured separately, and there is no consensus on the trade-offs between them.

Univariate Signals from E&S Data

In this section we address key issues arising in the constructing signals based on Environmental and Social data. One important theme is the need to normalize based on some proxy of activity or firm size. In this case study we decided to use revenue since it is something that applies uniformly to all companies. Another key theme is the need to make *relative* comparisons – the constructed signals reflect normalized performance relative to a peer group. This results in signals being normalized by both firm size and industry. But perhaps the most important theme developed here is that of estimating the marginal impact as companies scale. As we shall see, for many fields, there is strong statistical evidence that larger companies have a lower marginal impact than smaller companies – something that is missed by using pure intensity ratios and can bias signals if not accounted for.

Before getting in to the details, we emphasize that the techniques developed here have widespread applicability across various ES factors. In this study, GHG Scope 1, GHG Scope 2, Energy Consumption, Number and Amount of Spills and Fatalities are all fields that merit treatment as described below. Moreover, these techniques apply to ESG-related fields *not* in our study, such as Nitrogen Oxide Emissions (NOX), Volatile Organic Compounds Emissions (VOC), Sulphur Dioxide Emissions (SO2), Particulate Matter (PM10, PM5, etc.), Number and Amount of Fines (environmental and anti-competitive), Number of Environmental Incidents, Number and Duration of Strikes and Lockouts, Number and Units of Recalls, Number of Data Breaches and various others.

In the section below we treat emissions fields in detail. Fields such as Fatalities and Spills are count data and fields like Spill Amounts are heavily skewed due to a large number of zeros in the data. The treatment of count and zero-inflated data differs only because the dependent variable, Impact, is far from Gaussian in those cases. These present technical rather than conceptual challenges, and as such are detailed in the Appendix.

Emissions: Inferring diminishing marginal impacts

Consider the Environmental fields, GHG Scope 1 and Scope 2 Emissions[1], and Energy Consumption. It would be misleading to rank companies based on these numbers alone simply because they produce at different levels, and the emissions of small companies are expected to be smaller than emissions of larger companies. As some measure of efficiency is required, it is typical that these numbers are normalized by some measure

[1] Recall that Scope 2 GHG Emissions are indirect emissions, such as via energy use, and Scope 1 are the direct emissions resulting from business operations.

of activity, such as production levels or revenue. These are called *intensity ratios*, and measure environmental impact versus unit activity:

$$\text{Intensity} = \frac{\text{Impact}}{\text{Activity}}$$

Let's rewrite this as

$$\text{Impact} = \text{Intensity} \times \text{Activity}^{\gamma}$$

Here we have introduced the exponent γ to capture possible nonlinearity between production activities and environmental impacts. This formulation is consistent with the Cobb-Douglas production function, and γ captures the *elasticity* between production activity and environmental impacts. Note that

- If $\gamma=1$we have constant marginal impact, regardless of activity level
- If $\gamma<1$ the marginal impact is decreasing with activity
- If $\gamma>1$ the marginal impact is increasing with activity

The idea of examining environmental impacts in terms of elasticities and Cobb-Douglas like impact functions is not a new one. See, for example, (Ellis and Fisher) (Fullerton and Ta) and (York, Rosa and Dietz).

We can write this equation in logarithmic terms, and introduce an innovation term ε

$$\log \text{Impact} = \text{Average} \log \text{Intensity} + \gamma \times \log \text{Activity} + \varepsilon$$

This allows us to estimate the elasticity γ using regression techniques. Note that the company-specific term is now the innovation, and the intensity term is the intercept. If we estimate regressions by industry, then the intercept represents an average intensity for the industry, which may be useful for ranking industries. Furthermore, and most importantly, instead of assuming $\gamma=1$, we can estimate it, which potentially eliminates bias due to misspecification.

Plots of Scope 1 Emissions versus Revenue data are shown in Figure 3. On the left we see the data in nominal terms, whereas on the right we have used a logarithmic plot. Not only is it easier to see across the multiple orders of magnitude in the data, we can observe that the slopes of the various industries are indeed similar and near $\gamma=1$.

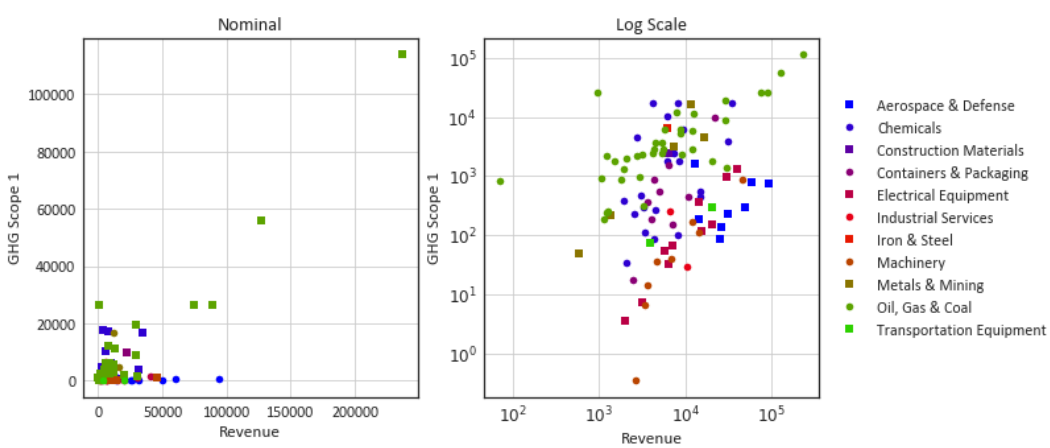

Figure 3: GHG Scope 1 Emissions versus Revenue, by industry. The plot on the left is in nominal coordinates, and the plot on the right is the same data graphed on a logarithmic scale.

In this case study we estimated the regression using a linear mixed-effects model, where the elasticity γ is a "fixed effect" across all industries, and the intercepts are "random effects" depending on industry. Results are shown in Table 4.

The results of regressing Scope 2 emissions on revenue are shown in Table 5. Here, the point estimate of elasticity is γ=0.81, which indicates a diminishing marginal impact. That is, for indirect GHG emissions, the data suggest larger companies might be in a position to take advantage of economies of scale more so than with their direct Scope 1 emissions. We also note that the industry ranks differ between Scope 1 and Scope 2.

Industry	Intercept	Unit Intensity
Oil, Gas & Coal	0.91	8.13
Iron & Steel	0.81	6.45
Metals & Mining	0.64	4.36
Construction Materials	0.54	3.46
Chemicals	0.47	2.95
Containers & Packaging	0.09	1.23
Transportation Equipment	-0.40	.398
Industrial Services	-0.56	.275
Aerospace & Defense	-0.64	.229
Electrical Equipment	-0.75	.178
Machinery	-1.10	.079

Table 4: Results of regressing log of Scope 1 emissions on log of Revenue using a linear-mixed effects model. The industry specific intercepts are shown above, and the estimated elasticity is 0.945 with a 95% confidence interval of [0.708, 1.18]. The industry unit intensities are 10^Intercept, with units of millions of metric tons.

The last step for signal extraction is transforming the innovation to a 0 to 1 scale. Using the normal cumulative distribution function Φ, preserves percentiles and converts large innovations to either zero or one depending on sign. For emissions, negative innovations are "good", so we use

$$1 - \Phi(\varepsilon_i; 0, \sigma)$$

where σ is the standard deviation of the innovations. Effectively this makes the signal equal to the percentile of size-normalized emissions except that we have captured elasticity effects.

Industry	Intercept	Unit Intensity
Iron & Steel	0.55	3.59
Oil, Gas & Coal	0.40	2.50
Containers & Packaging	0.33	2.13
Chemicals	0.24	1.74
Construction Materials	0.24	1.74
Metals & Mining	0.07	1.18
Transportation Equipment	-0.13	0.75
Aerospace & Defense	-0.20	0.63
Electrical Equipment	-0.25	0.57
Industrial Services	-0.46	0.34
Machinery	-0.80	0.16

Table 5: Results of regressing log of Scope 2 emissions on log of Revenue using a linear-mixed effects model. The industry specific intercepts are shown above, and the estimated elasticity is 0.806 with a 95% confidence interval of [0.565, 1.05]. The industry unit intensities are 10^Intercept, with units of thousands of metric tons.

Results of the signal extraction model are illustrated in Figure 4. We point out that the traditional intensity approach introduces a bias that can benefit large companies and can penalize small companies. This is because innovations assume unit elasticity, so large companies behaving in a linear way are not penalized. A few company-specific examples are shown in Table 6. Two of the largest companies, Phillips 66 and Marathon, have better-than-average intensities, but their actual Scope 2 emissions are greater than the expected value based on regression. Conoco Phillips, on the other hand, has revenue of a similar magnitude, and has Scope 2 emissions below the diminishing impact expected value. Naturally it fares well under both methods, but signal method developed here better recognizes their diminishing marginal impact. Conversely, Diamondback Energy has an order of magnitude smaller revenue, is in the bottom half based on intensity, but is in the top half based using the signal constructed here, reflecting a more careful comparison to its revenue peers

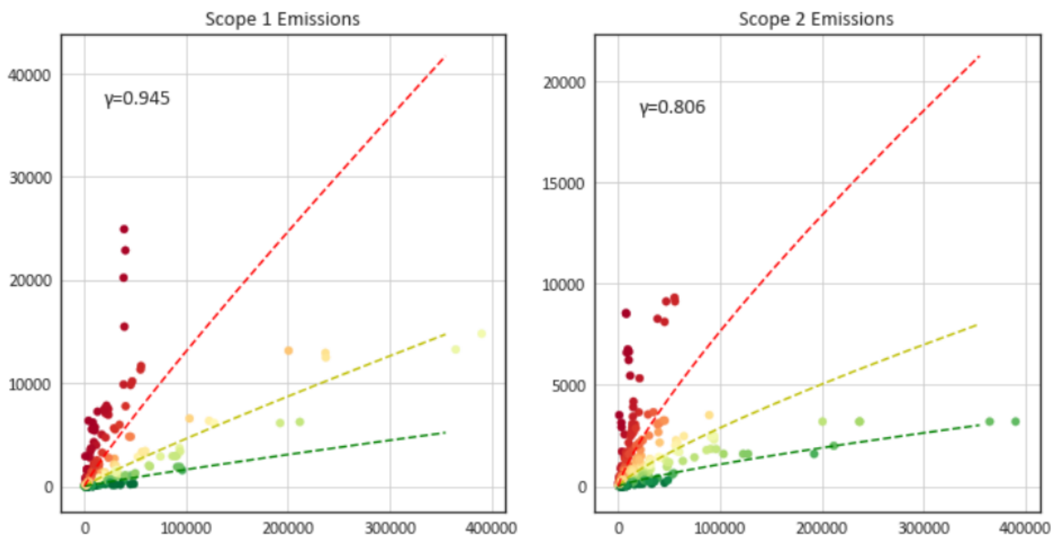

Figure 4: Signal extraction results for Scope 1 and Scope 2 emissions. The color of the dot indicates the signal value, based on the innovation, with green being near 1 and red near 0. The dependent variable (vertical axis) is graphed with the industry-specific intercept acting as a normalization constant. The dotted lines represent the mean and plus/minus one standard deviations based on the regression estimates.

Finally we point out that when $\gamma \approx 1$, the resulting signal is comparable to using intensities. In fact, when $\gamma = 1$ and if two intensities are the same, the innovation will be the same[2].

Company	Revenue	Scope 2	Intensity	Intensity Half	Expected Scope 2	Signal	Signal Half
Phillips 66	$89,300	8,800	0.10	Top	6,723	0.41	Bottom
Marathon	$74,733	7,600	0.10	Top	5,818	0.41	Bottom
ONEOK	$12,174	2,100	0.17	Bottom	1,333	0.35	Bottom
ConocoPhillips	$29,106	1,200	0.04	Top	2,705	0.75	Top
Diamondback Energy	$1,212	158	0.13	Bottom	205	0.59	Top

Table 6: Selected differences between regression-based signals and intensities. Large companies rank better by intensity than they do based on expected values via regression. Conversely, small companies can fare worse with intensities. Both are evidence of bias in using intensities when data exhibit diminishing marginal impacts.

[2] To see this, note that two companies have the same intensity if both the impact and activity are scaled by the same amount, say m. Then

$$\varepsilon = \log m \times \text{Impact} - \log \text{Intensity} - \log m \times \text{Activity} = \log \text{Impact} - \log \text{Intensity} - \log \text{Activity}$$

Health and Safety Rates

ESG frameworks typically consider Human Capital in the Social Dimension, and health and safety rates are used as a quantifiable measure of outcomes for this. For the purposes of this case study we assume the investor's mandate includes the view that human capital should be treated equally in all industries and regardless of whether the employee is full-time or a contractor. This implies that we estimate signal curves across the entire population. With this assumption, there is no industry specific peer group, and signal averages may be different for different industries.

As discussed previously, if the investing mandate includes limited sector-driven tracking error, one could justify using an industry normalization, effectively implementing an ESG "tilt".

Incident Rates. In the US, safety incident rates are defined by and reported to the Occupational Safety and Health Administration agency (OSHA). For this study we consider two commonly used metrics to assess worker safety:

- *Lost Time Incident Rates (LTIR)*: any occupational injury or illness resulting in the employee being unable to work a full assigned shift.
- *Total Recordable Incident Rates (TRIR)*: any occupational injury or illness requiring medical treatment or first-aid.

Since that Lost Time incidents are included in the Total, we have TRIR > LTIR.

Since only the rates are available to us (and not numerator and denominators), we are essentially forced to assume unit elasticity versus aggregate hours worked in a company. Because the data are non-negative and may have a mode greater than zero, we use the gamma distribution. The survival function (one minus CDF) is used so that the signal is 1 when incident rates are zero, and declines reflecting the data percentiles. The data distributions and resulting signal curves are shown in Figure 5.

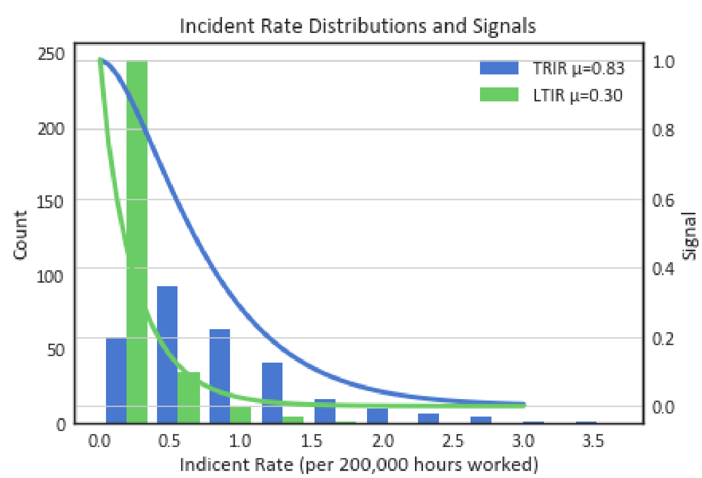

Figure 5: Histograms of Total Reportable Incident Rates and Lost Time Incident Rates (left axis) and signal curves (right axis).

Fatalities. These are normalized based on workforce size. The treatment is similar to emissions except that

- The data has far more zeros than can be explained with a (transformed) normal distribution. Therefore we employ a modified regression method than can account for skewed dependent variables, which is discussed in the Appendix.
- As with Incident Rates, we assume a mandate that does not differentiate by industry.

Community Rights & Relations Policies (CR&R)

Our case study includes two indicator variables

- *Human Rights Policy*. Indicates whether the company has implemented any initiatives to ensure the protection of the rights of all people it works with. “False” values indicate that the company has not explicitly disclosed any such efforts in its most recent Annual or Company Responsibility reports.
- *Indigenous Rights Policy.* Indicates whether the company has disclosed a management approach for managing relations with indigenous populations in areas where they operate. The policy may include community relations, stakeholder engagement, local employment, local social and economic development, feedback and grievance mechanisms, land and resource use and resettlements before during and after mine operations.

For these variables, we set the signal values to be 0 or 1 accordingly. Admittedly this is a blunt measure, as we do not consider the content or strength of these policies (which would require sentiment analysis, perhaps via NLP).

Univariate Signals from Governance Data

Governance data differs from ES data in a few key ways. First of all, it is not subject to company size normalizations, at least not in the same fundamental way. Secondly, corporate governance is closely aligned with traditional investment objectives, and as such there is a combination of guidelines, accepted best practices and even regulations that can guide designing signals that measure good governanc[3]. Finally, a key difference between governance fields and ES fields is that signal curves may not be strictly increasing or decreasing as a function of the data.

With the exception of Gender Diversity, in this case study we demonstrate a technique for extracting signal shapes based on predicting financial and environmental performance from historical data. To be sure, this is subjective decision, as one goal of this paper is to demonstrate a variety of possible approaches without claiming optimality. What practitioners ultimately decide to do should certainly depend on their own mandates and objectives.

Board Gender Diversity

There is now widespread demand for improved gender diversity in corporate board rooms. Major asset managers and banks have put companies on notice that all-male boards are unacceptable and that they will be held accountable for improvement.

Our task here is to *quantify* this sentiment by assigning the percentage of women directors a signal value. To this end we refer to the Bloomberg Gender Equality Index, which for 2019 exhibits an average 26% women directors. This is our double the 13.3% sample average of women directors in our data set. Our admittedly subjective approach is to assign a signal value of 0.5 to the average and 0.8 to the GEI value of 26%. Of course 50% is assigned a signal value of 1[4]. The results of this approach are shown in Figure 6.Notably for 2017, 40 out of 260 companies have no women board members, leading to a substantial proportion of companies receiving a signal value of 0.2 or less.

---

[3] For example, see the ISS US Proxy Voting Guidelines.

[4] We are not concerned with defining signal values beyond 50% (where at some point it might conceivably decrease) since our sample maximum is 50% and most boards today are far from 50%.

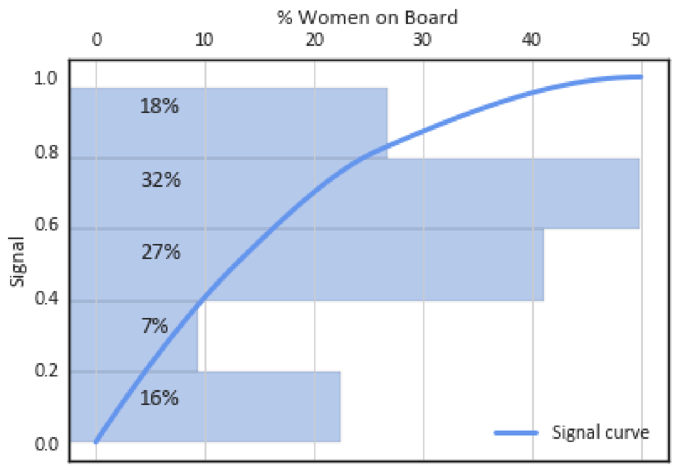

Figure 6: Signal curve and resulting distribution for percent women directors for 2017. The numbers in each bar are proportion of companies in each signal quintile.

We emphasize that we have relied on principles here – other than the introduction of an interpolating curve, there is no model or estimation involved. Why not use data? While it would be possible to do so, we note that historically boards have been nearly all male. This can make it difficult to establish a pattern from the thin sub-sample of firms where women have significant representation in the boardroom, as there are many other factors to control for. Moreover, besides historically low representation, the benefits of improved diversity are likely to be occur over several years, and possibly entire economic cycles, which would require a different historical data set than this case study covers. Ultimately, however, our choice demonstrates how one might incorporate a kind of external guideline that is not unusual to sustainable investing mandates.

Board Experience, Refreshment and Independence

We now return to a more data-driven approach, first observing that we have to address the fact that governance data may not be monotonically "good" or "bad". That is, we likely want to assign high signal values to intermediate values reflecting a "sweet spot".

SIGNALS	AGE RANGE	BOD AVERAGE AGE	CHAIRMAN TENURE	NUM BRD MEMB SERVING OVER 5Y	NUM BRD MEMB SERVING OVER 10Y	PCT INDEPENDENT DIRECTORS
GHG SCOPE 1			0.11	-0.12	-0.07	-0.18
GHG SCOPE 2			0.10	-0.12	-0.08	-0.19
ENERGY CONSUMPTION		-0.06		-0.12	-0.11	-0.19
NUMBER SPILLS				0.06		
AMOUNT OF SPILLS		-0.08				
LTIR		-0.07	0.08	-0.16	-0.13	-0.17
TRIR		-0.06	0.10	-0.10	-0.08	-0.15
HUMAN RIGHTS POLICY	-0.08		-0.12		-0.06	0.23
EXCESS RETURN				0.09	0.06	0.07
VOLATILITY		-0.07		-0.20	-0.18	-0.10

Table 7: Spearman rank correlations between E and S signals and Governance fields. Only correlations with p-value less than 0.05 are shown. Yellow cells indicate that younger or less tenured directors are positively associated to the signal, while blue cells indicate older or longer tenured directors are associated with higher signal values. There is no clear pattern for the percentage of independent directors. There were no significant correlations to Fatalities.

As initial motivation for such non-monotonic relationships, Table 7 contains the Spearman rank correlations between various governance metrics and quantities of interest, including the Environmental and Social signals already constructed and the one-period ahead financial performance captured as Excess Return (defined as the company stock return minus the S&P 500 Ex-Financials and Real Estate index) and volatility. While the correlations are modest, the reported results are all statistically significant with p-values of 0.05 or less. As the table shows, the results are mixed: some governance fields are positively correlated (blue) to key indicators while others are negative (yellow)[5]. Further evidence for non-monotonicity can be observed via boxplots as shown in Figure 7. While the variation from one quintile to the next is not statistically significant, one can see that the lowest quintile volatilities occur somewhere in the middle of the distribution.

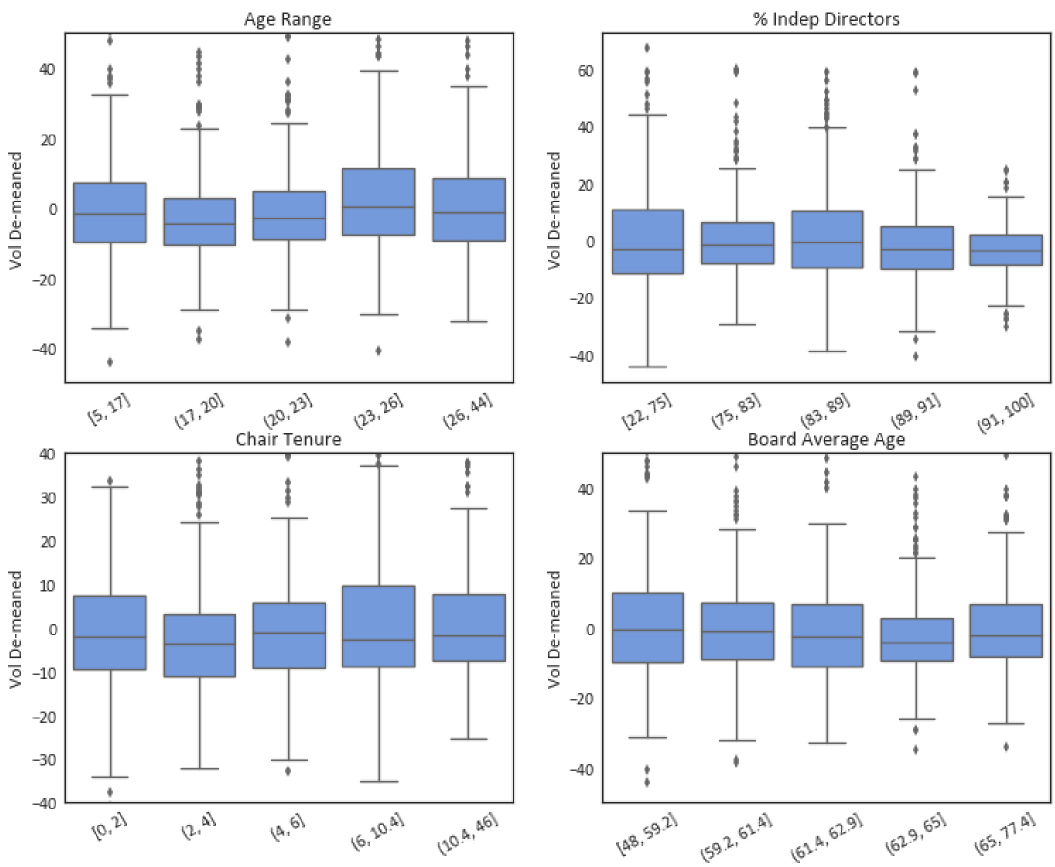

Figure 7: Box plots of one-period ahead demeaned volatility by quintiles of governance field values. Assuming lower volatility is the objective, one would design the signal to peak around the quintile with the lowest mean.

Our strategy is to look for a signal function $F(X, \theta)$, where X is the original univariate governance data and θ are unknown parameters that control the signal shape. This signal function should be useful for predicting various outcomes, such as excess returns, volatility or Environmental and Social performance. This set-up amounts to the non-linear multivariate regression

$$Y = c + \beta F(X, \theta) + \varepsilon$$

Such regressions can be estimated numerically using non-linear optimization or MCMC routines. Key details and results of our estimations are provided in Table 8.

[5] Except for volatility, which is reversed and colored blue because lower volatility is the objective.

In spite of the data-driven approach, some decisions here are subjective. We first examined box plots such as those in Figure 7 – raw metric quintiles versus target variables. Preference was given to cases exhibiting a single mode between governance signals and financial performance (improved excess returns, reduced volatility). When a similar mode appeared versus carbon emission signals, those signals were also included as dependent variables for estimation robustness. If the box plots did not exhibit evidence for a single mode, the target was eliminated from the MCMC optimization. This process led to different targets being used in the search for signal shape for different governance fields.

Additionally, the MCMC estimation for Board Average Age is the most sensitive to the prior (initial value) and has a wide confidence interval, suggesting that its signal shape was not well-learned from the data. The remaining signals are statistically stronger, but the results do reflect choices of shape family and prior. Thus one might view this approach as a hybrid between subjective and objectively optimal parameter settings.

G Metric	Motivation	Targets	Treatment
Board Age Range	Age range captures an aspect of diversity, which can encompass experience and multiple investment horizons.	Returns, Volatility, Energy Use	Gamma curve with mode at 22.7 years, 95% CI of [21, 31.4]. Prior/initial value = 19.
Board Average Age	A higher average age may be correlated with both experience and entrenchment.	Volatility	Normal curve with mode at 59.97 years, 95% CI of [49.7, 73.3]. Prior is 60.
Chair Tenure	Long chair tenures are associated with entrenchment and perceived to be detrimental for corporate governance. However, short tenures may be associated with inexperience.	Volatility, Energy Use	Gamma curve with mode at 6.8 years, 95% CI of [-0.4, 15.4]
Board Tenure	Existing research suggests that regular refreshment, resulting in balanced board tenures results in better returns with lower risk and that, conversely, boards with concentrations in either low or high tenure buckets have inferior performance on average.	Returns, Volatility, Carbon Emissions, Energy Use	Dirichlet surface with mode at 25% directors in 0-5Y tenure bucket, 20% in 5-10Y bucket and 55% in 10+ bucket. CIs are [.18, .31], [.06, .27], and [.48, .72]. Prior/initial value = [1/3,1/3,1/3].
Percentage of Independent Directors	Governance best practices include a strong role for independent directors to ensure that management is held accountable to the broader company, including shareholders.	Volatility, Carbon Emissions	Beta curve with mode at 81%, 95% CI of [0.76, 0.84]. Prior/initial value = 0.75

Table 8: Summary of governance signal definitions.

Results are shown in Figure 8. Note that Age Range and Board Average Age are far from percentile-like signals: 73% and 56%, respectively, of those signals are clustered into signals 0.8 or higher. One interpretation of this is that signal is effectively acting as indicators of poor performance, and something that the majority of companies "pass".

Figure 9 shows analogous results for the distribution of board tenures using a ternary plot. Each of the three axes represents the percentage of directors in each tenure bucket. Since the percentage must sum to 100%, the plot is shows the triangle formed by the graph of the plane x+y+z=1 in the positive octant of a 3D graph. High signals are in the lower left because the estimated result has 55% of directors in the 10+ buckets.

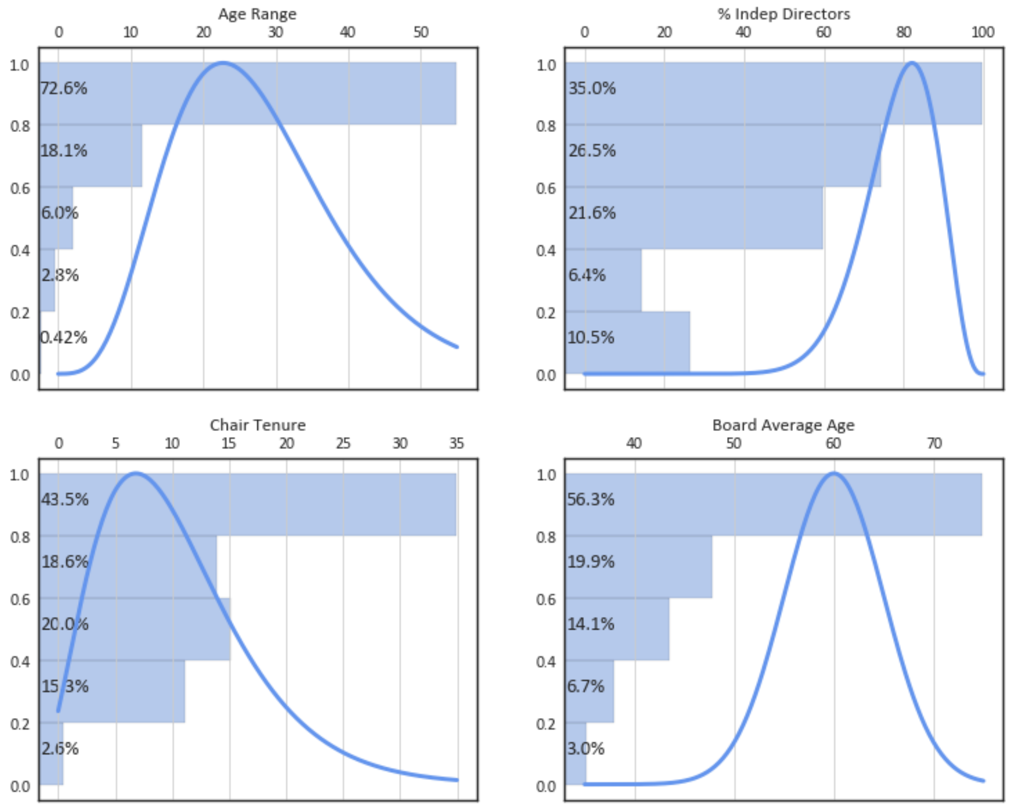

Figure 8: G signal curves and signal distributions. Horizontal bars depict the distribution of resulting 0-1 signals. The dark blue lines are the signal curve.

A final note on the selection of non-monotonic signal shapes: an initial version of this case study involved only monotonic shapes based on the percentile of the observation – considered by many to be "safe ground" for ESG ranking. In that case the resulting aggregated G signal exhibited no significant relationship with either one-period ahead returns or volatility. As we shall see, the outcome is different with signals constructed as above.

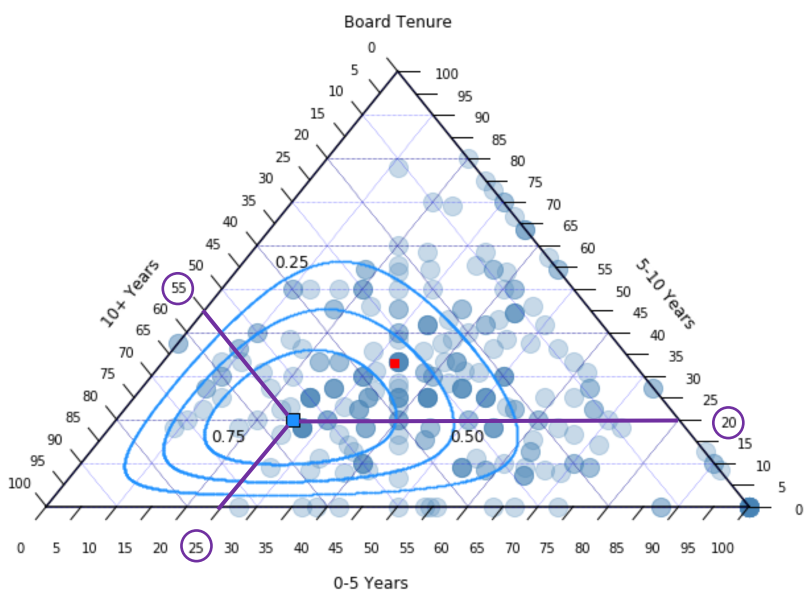

Figure 9: Signal levels for board tenure. Blue contour lines are the level curves for signals 0.25, 0.50 and 0.75, with the blue square set at the peak signal mode (25%, 20%,55%). The red square is the location of the uniform tenure distribution of 1/3 in each bucket and receives a signal of 0.625. Gray-blue dots represent the density of observations.

Signal Aggregation

We now turn our attention to combining or aggregating signals. As with the previous sections, there are both objective and subjective aspects to our approach. Depending on one's objectives or mandate, different decisions could certainly be justified.

We begin with factor analysis, which can be used to determine if first-layer signals are statistically redundant in that they differ only by noise. In such case it makes sense to average those signals in the hope of noise cancellation.

Outside of those cases, we must find an approach to rank companies based on (or in spite of) multivariate characteristics. This cannot be done without expressing some beliefs about the relative importance and tradeoffs among individual characteristics.

Factor Analysis

We begin by separately exploring the correlations within the ES and G signal groups. Table 9 shows the correlations among the ES signals, where we have ordered the fields so that the signals for Spills, Carbon, CR&R, Fatalities and Health & Safety are adjacent.

	NUMBER SPILLS	*AMT SPILLS*	*GHG SCOPE 1*	*GHG SCOPE 2*	*ENERGY CONSUMPTION*	*IR POLICY*	*HR POLICY*	*FATALs EMP*	*FATALs CNTRCTR*	*TRIR*	*TRIR CNTRCTR*	*LTIR*	*LLTIR CNTRCTR*
NUMBER SPILLS	1												
AMT SPILLS	0.77	1											
GHG SCOPE 1	0.18	0.12	1										
GHG SCOPE 2			0.88	1									
ENERGY CONSUMPTION	0.26	0.21	0.59	0.51	1								
IR POLICY	0.31	0.37				1							
HR POLICY		0.13	-0.28	-0.26	-0.35	0.34	1						
FATALs EMP	0.54	0.51	0.40	0.32	0.44	0.20		1					
FATALs CNTRCTR	0.61	0.71			0.21	0.24		0.51	1				
TRIR	0.15		0.51	0.54	0.44		-0.31	0.33		1			
TRIR CNTRCTR	0.71	0.69	0.24	0.13	0.24	0.38		0.59	0.64	0.16	1		
LTIR	0.36	0.25	0.55	0.45	0.53		-0.25	0.44	0.21	0.53	0.31	1	
LTIR CNTRCTR	0.70	0.69	0.20		0.24	0.29		0.55	0.75		0.86	0.34	1

Table 9: Spearman rank correlations between univariate ES signals. Only correlations with a p-value less than 0.05 are shown. The highlighted block diagonals show the in-group correlations among Spills, Carbon, CR&R, Fatalities and Health & Safety.

The relatively high correlations along the block diagonals are candidates for factor analysis. Indeed, one can calculate factor loadings for the fields in each block to seek further evidence that these groups are single-factor. Results are shown in Table 10. While the traditional threshold of 0.7 is not a rigorous guideline, the result is consistent with our starting ESG framework.

Signal Group	Univariate Signals	Factor Loadings
Spills	Number Spills, Amt Spills	**1.58**, 0.42
Carbon	Scope 1, Scope 2, Energy Consumption	**2.15**, 0.62, 0.23
CR&R	Human and Indigenous Rights Policies	**1.45**, 0.55
Fatalities	Fatalities Employees, Fatalities Contractors	**1.0**, 0.0
Health & Safety	TRIR+LTIR, Employees and Contractors	**2.97**, 0.62, 0.32, 0..08

Table 10: Factor loadings for each signal group. Based on the Kaiser criterion of 0.7, each group has only one significant factor (in bold).

As a result of this factor analysis, we construct group signals for Spills, Carbon, CR&R, Fatalities and Health & Safety by taking simple averages of the univariate signals. Note there is also a missing data benefit here: if a univariate signal is missing, we average what's available without penalty. Statistically, missing data within such groups can be considered more as a missed opportunity to reduce noise than truly missing a signal.

A similar analysis for G signals shows less evidence for dimension reduction. Statistically significant rank correlations are shown in Table 11, and the factor loadings are shown in Figure 10. Moreover, since there is very little missing data for each G signal, we will refrain from aggregating using a simple average and instead look for principle and evidence-based methods for combining into an overall G signal.

	Pct Indep Dirs	*Chair Tenure*	*Board Avg Age*	*Age Range*	*Board Tenure*	*Pct Women on Board*
Pct Indep Dirs	1					
Chair Tenure	0.16	1				
Board Avg Age		0.13	1			
Age Range				1		
Board Tenure				0.24	1	
Pct Women on Board					0.13	1

Table 11: Spearman rank correlations for governance signals. Only correlations with a p-value greater than 0.05 are shown. Visually the correlations can be arranged into the two blocks shown, but it is unclear what, if any, interpretation to give to them.

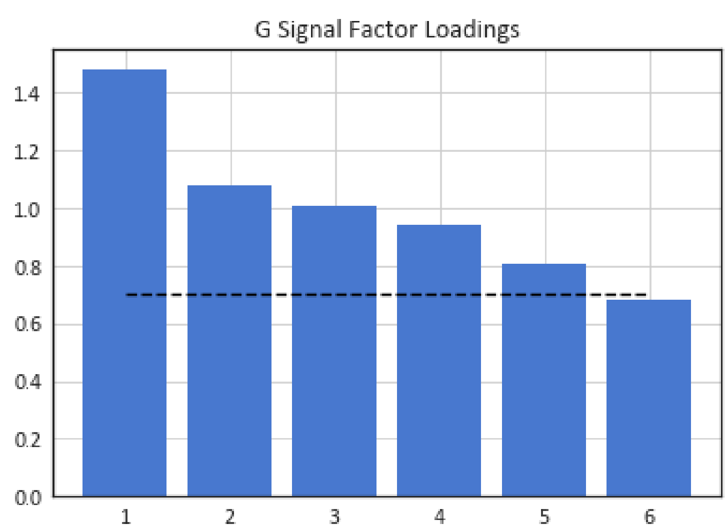

Figure 10: Factor loadings for G signals. Based on the Kaiser criterion of 0.7, there are at least 5 factors among these 6 signals.

Composite Indicators and Generalized Means

Aggregating multivariate signals into a single "overall" signal is both discomforting and familiar. On one hand, doing so obscures information about empirically different features of the entities being ranked. On the other, in everyday life we have generally come to value how complex traits can be summarized via ranking algorithms, viz. college rankings, all-time greatest sports rankings, economic rankings, etc.

Indeed the OECD has founded the Competence Centre on Composite Indicators, which sponsors research and collects best practices for aggregating metrics and signals into overall indicators that are suitable for policy decisions. One of the key design decisions they highlight concerns *compensability*, meaning the extent to which performance in one signal can offset poor performance in another. When two traits are reasonably substitutable, it makes sense to allow compensability in signal aggregation. In other cases, it might be reasonable to expect good performance in multiple signals. Note that the introduction of weights into a linear average does not fundamentally alter substitutability – weights merely define the *rate* of substitutability.

One way to implement limited compensability is to use Generalized Means. Given a vector of n signals s, a weight vector w and an exponent p, the Generalized Mean is given by

$$M(s, w, p) = \left(\sum_{j=1}^{n} w_j s_j^p \right)^{1/p}$$

This formula encapsulates some well-known special cases:

p = -∞	Minimum
p = -1	Harmonic mean
p = 0	Geometric mean
p = 1	Arithmetic (simple) mean
p = ∞	Maximum

Table 12: Special cases of the Generalized p-Mean.

Qualitatively one can think of the choice of p in terms of the following:

- When p<1, there is a *penalty* for uneven performance – the result is skewed toward the lowest value. One might set p<1 when it is reasonable to expect all input signals to be high – perhaps when there are clearly defined quantitative signals and no substantial trade-offs in trying to improve them.
- When p>1, the aggregate signal provides a kind of *benefit of doubt* – the result is skewed towards the highest value (best input signal). Choices of p>1 can make sense in cases where there are multiple ways to achieve a goal. For example sufficient Board Independence might be met either through independent directors or independent leadership.

The choice of parameter p may be best explained by example. Consider ranking three hypothetical companies based signals for Diversity, Refreshment and Independence as shown in Table 13.

- Company A and B have the same average (simple mean) – 0.7
- Company A has perfectly even performance while Company B is very uneven.

- Meanwhile Company C has a lower average performance, but is a relatively even performer.

This situation poses some important questions:

- Are companies A and B "the same"?
- Is company B really better than C?

	Company A	Company B	Company C
Diversity	.7	.1	.4
Refreshment	.7	1.0	.6
Independence	.7	1.0	.8
Harmonic Mean (p=-1)	.7	.25	.55
Geometric Mean (p=0)	.7	.46	.57
¼ Mean (p=1/4)	.7	.53	.58
½ Mean (p=1/2)	.7	.6	.59
Simple Mean (p=1)	.7	.7	.6

Table 13: Hypothetical cases for aggregating Diversity, Refreshment and Independence signals.

The views one has for those questions can provide guidance on selecting p, as any p<1 will result in company B receiving a lower aggregate than A, and a choice of p<½ results in company C receiving a higher aggregate signal than B. In short, the choice of p is a parsimonious way to implement a view on even performance. Weights can be used if one has a view on relative importance independently from even performance.

Aggregate G Signal

Although relying on judgement could be one approach for parameter selection, we now demonstrate how one can use historical data to relate parameters to financial performance. This is possible with G data because of the relatively long history. As with the univariate G signals, we can pose the problem as a non-linear regression where the weights w and exponent p are unknown.

$$Y_j = C_{i(j)} + \beta M_\delta(s_j, w, p) + \varepsilon$$

For the dependent variables Y, we use one period ahead volatility and excess returns. Note we included industry-specific intercepts. In addition, we incorporated a shift term δ to avoid possible singularities for zero signals[6]. The main results of the estimation are shown in Table 14.

The estimation is initialized with a prior of equal weights and results in only modest differences from that: based on the data, Independence becomes slightly down-weighted and Percent of Women on Board becomes slightly over-weighted. The prior for p was set to 0.5, and the result is close to that value as well – although we note that the confidence interval for p is strictly less than 1.0, giving evidence for limited compensability.

Visualizations of the regression results are shown in Figure 11. We also observe that while the slope for returns is not statistically significant, the estimated G signal does correlate with lower volatility. The estimated β is -11.1 and is statistically significant, indicating that as the G signal moves from 0 to 1, the excess volatility is expected to

[6] To be precise, the signal is shifted and the mean is un-shifted: $M_\delta = M(s+\delta, w, p) - \delta$. In this study we fixed δ=0.5.

decrease by 11.1%. The rank correlation between the aggregate G signal and volatility is -0.20.

Parameter	Mean	CI (95%)
p	0.48	[0.1,0.87]
β_{vol}	-10.3	[-16.9, -3.6]
β_{return}	2.5	[-8.8, 13.7]
w: Pct Indep Dirs	0.126	[0.07, 0.25]
w: Chair Tenure	0.161	[0.07, 0.26]
w: Board Avg Age	0.163	[0.04, 0.25]
w: Age Range	0.170	[0.05, 0.22]
w: Board Tenure	0.165	[0.07, 0.19]
w: Pct Women on Board	0.215	[0.11, 0.32]

Table 14: Parameter estimates of for G signal aggregation via generalized mean. Note the β for returns is not statistically significant since zero is contained in the CI.

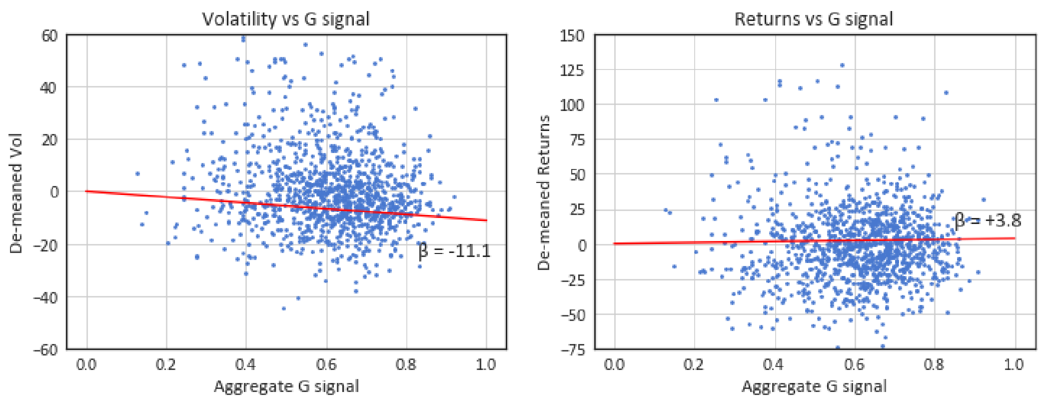

Figure 11: Regression of aggregate G signal on volatility and return. The dependent variables are demeaned by industry and year. Keep in mind the slope for returns is not statistically significant.

One can also examine the relationship between one-period ahead volatilities and the G signal non-parametrically by quintiles, as shown in Table 15. Although the signals were constructed using in-sample data, it is encouraging that higher G signals are consistent with lower volatility in each individual year.

Year	G Q1	G Q2	G Q3	G Q4	G Q5
2013	2.9	3.4	-2.8	2.0	-3.8
2014	8.3	1.0	-3.6	-4.1	-1.8
2015	4.2	0.0	-0.9	-2.5	-0.6
2016	-0.7	3.6	0.5	0.8	-4.0
2017	3.8	-0.5	0.1	-0.6	-3.4

Table 15: G Signal quintiles of one-period ahead volatilities demeaned by industry and year.

Aggregate ES Signal

Combining the ES group signals into an aggregate signal is more difficult due to the lack of historical data, although in practice we are not terribly sensitive to whatever we chose because few industries have much of a footprint in more than 1-2 groups. Table 16 displays the percent of tickers for which a univariate signal is available in the 2017 reporting year. Carbon is the best represented, with data completeness highest in Containers & Packaging, and lowest in Oil, Gas & Coal, with data available for only 11 out of 99 companies.

We therefore fix the weights by considering signal availability, its rank correlation to one-period ahead volatility[7] and signal type. For example, CR&R is down-weighted to 1/8 since it is based only on indicator variables. Fatalities and Safety receive the same weight due to low availability. The remaining weight ratio of 1.5:1 between Carbon and Spills reflects the better availability of Carbon and Energy data used in its construction.

We also have a minimum threshold: there must be at least one non-policy based signal available (i.e. one signal besides CR&R), otherwise the aggregate ES signal is set to null for that ticker and reporting year. Finally we chose p=1/2, based both on the view that compensability is not readily justifiable, as well as for consistency with the G signal.

		Signal Availability				
Industry	**Tickers**	**Carbon**	**Spills**	**Fatalities**	**Safety**	**CR&R**
Aerospace & Defense	18	33%	-	-	-	6%
Chemicals	33	58%	-	3%	15%	-
Construction Materials	8	13%	13%	-	13%	-
Containers & Packaging	12	67%	-	-	-	-
Electrical Equipment	24	25%	-	-	-	4%
Industrial Services	8	13%	-	-	-	-
Iron & Steel	10	20%	-	-	-	70%
Machinery	29	17%	3%	-	-	-
Metals & Mining	11	27%	9%	27%	9%	100%
Oil, Gas & Coal	99	11%	21%	2%	11%	99%
Transportation Equipment	10	20%	-	-	-	-
Vol Rank Correlation		**-0.16**	**-0.2**	**-**	**-**	**-0.15**
Assumed Weight		**3/8**	**1/4**	**1/8**	**1/8**	**1/8**

Table 16: Signal availability per Industry for 2017. The bottom row indicates our resulting choice for weight based on reporting and correlation to excess volatility.

As a result of these assumptions, an aggregate ES signal can only be calculated for 50-85 tickers per year (out of 280). With the important caveat that this may not be representative, the rank correlation between the aggregate ES signal and one-period ahead volatility is -0.23 with a p-value of 10^{-5} – greater than any of the rank correlations between the input signals and volatility. Furthermore, the β for regressing one-period ahead volatility on this ES signal is -7.8, and is statistically significant.

Aggregate ESG Signal

Given that low ES signal availability will have a strong influence on the results, it is not clear there will be much benefit in calculating an aggregate ESG signal. Nonetheless for completeness we present the results using equal weights between the ES and G signals, and p=1/2. The estimated β is -13.5 and rank correlation is -0.26, both marginally better

[7] The rank correlations between univariate ES signals and one-period ahead returns are all statistically insignificant, so we exclude that from consideration.

than the results of ES and G alone. This result is visualized in Figure 12.

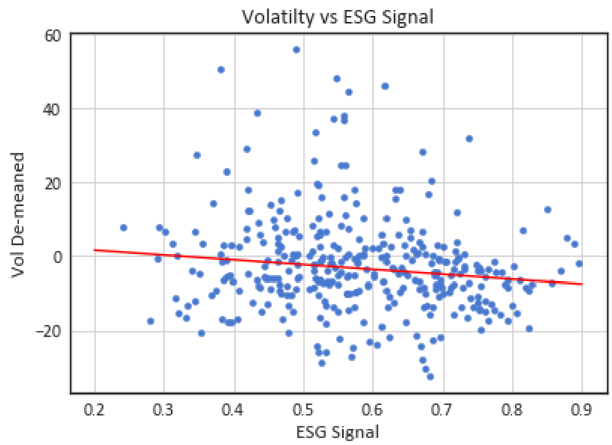

Figure 12: Overall ESG signal versus volatility, where the volatility is demeaned by industry and year. Here β=-13.0.

Discussion of Case Study Results

Cross Sections and Trends

Let us turn to a few other results besides the negative (in-sample) correlation to volatility. First, examining the histograms of the aggregate signals in Figure 13, we see that both avoid too much concentration. Having a reasonably "spread out" distribution suggests that we have preserved some ability to differentiate performance. Secondly, and more interestingly, we see differences in the mean and skew of the two aggregate signals. The ES signal is centered at 0.5, and has a skew of 0.06. This should be expected given that most of the univariate ES signals are derived from regression innovations.

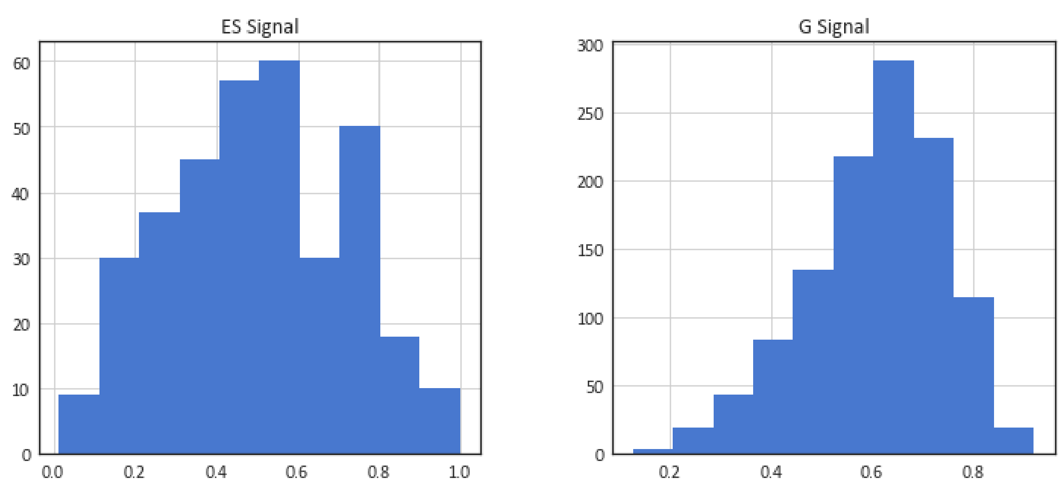

Figure 13: Histograms of aggregate signals.

On the other hand, the G signal has a mean of 0.61 and skew of -0.58. This may reflect a consequence of using financial performance variables in the signal design. Existing governance principles may compel the majority companies to have similar characteristics, including return characteristics, leading to outliers in both board composition metrics and financial performance.

In terms of trends, the results are modest. G signals show a slight increase over the 2013-2017 sample period, which is probably not statistically significant. ES exhibits a slight decrease, due to a decrease in the Spills signal from 2014 onwards. Such effects could simply be due to better reporting.

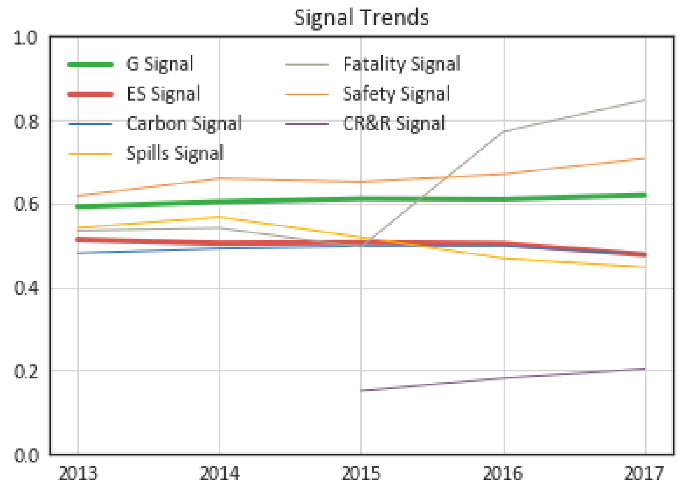

Figure 14: Year-over-year signal trends (cross-sectional means).

Company Analysis

In this section we examine a few company-level results to see the impact of our decisions, and to see if we can perhaps determine any common characteristics among them. There is a major caveat in trying to draw conclusions here: our case study has utilized only a subset of fields that many ESG practitioners consider to be relevant in order to highlight methodological issues. (Refer again to Figure 1 for some of the factors not included in our results.) In short, while we can examine the impact of our design decisions, we caution against making over-arching ESG conclusions based on this signal set.

Company	BICS1	BICS2	Revenue	2018 Vol	G Signal	Board Tenure	Age Range	Chair Tenure	Pct Indep	Brd Avg Age	Pct Women
Lennox International	Industrials	Electrical Equipment	3.8	-7.1	0.89	0.97	1	0.98	0.6	0.99	0.85
Regal Beloit	Industrials	Machinery	3.4	-9.0	0.88	0.71	0.97	0.99	0.98	0.98	0.68
WESCO International	Industrials	Industrial Services	7.7	-4.1	0.86	0.97	0.64	1	0.81	1	0.79
Avery Dennison	Materials	Chemicals	6.6	-4.7	0.85	0.99	1	0.99	0.65	1	0.64
Chevron	Energy	Oil, Gas & Coal	127.5	-15.8	0.85	0.18	0.99	0.98	0.99	0.94	0.92
Schlumberger	Energy	Oil, Gas & Coal	30.4	-20.3	0.84	0.6	0.97	0.7	0.99	0.98	0.79
General Dynamics	Industrials	Aerospace & Defense	31.0	7.6	0.83	0.61	1	0.94	1	0.61	0.82
HB Fuller	Materials	Chemicals	2.3	9.4	0.82	0.8	1	0.81	0.7	0.91	0.73
Cummins	Industrials	Transportation Equipment	20.4	-3.6	0.82	0.34	0.84	0.99	0.99	0.92	0.79
Helmerich & Payne	Energy	Oil, Gas & Coal	1.8	6.9	0.82	0.88	0.96	0.99	0.98	0.92	0.4

Table 17: Companies with highest levels of G signal for 2017. The 2018 volatility is demeaned by industry.

Table 17 contains the G signal and supporting data for the 10 companies receiving the highest aggregate G signal levels. There are 7 industries and two orders of magnitude in revenue represented in this list. While no company received an aggregate G signal 0.9 or higher, these companies tend to have high values for all but one input signal. Board Tenure has the lowest average performance among these 10, and 5 of the companies have input signals less than 0.75. The average for Percent Women on Board for this group is 0.74, although only one company, Chevron, has a signal value above 0.9. There are 3 companies receiving a signal of 0.75 or less for Percent Independent Directors.

In order to better understand which G signals drive aggregate performance, we show the average signal values for each quintile in Table 18. Age Range shows the least differentiation, which is consistent with the histogram of these signals, where 73% of companies have signal values 0.8 or higher. Board Average Age has the second least differentiation, which is also expected based on the signal histogram. On the other hand, Board Tenure and Percent Women on the board seem to be driving most of the variation, with Independence being somewhere in the middle.

	Aggregate G	Board Tenure	Age Range	Chair Tenure	Pct Indep	Board Avg Age	Pct Women
Q1	0.41	0.0	0.86	0.59	0.20	0.64	0.0
Q2	0.58	0.01	0.93	0.58	0.60	0.82	0.43
Q3	0.64	0.20	0.90	0.69	0.70	0.89	0.60
Q4	0.70	0.49	0.95	0.82	0.65	0.89	0.66
Q5	0.77	0.71	0.98	0.89	0.81	0.91	0.73

Table 18: Average signal values for each quintile of the 2017 G signal.

Recall that the motivation for using generalized means is consistent performance across ESG factors. In Table 19 we calculate the difference between using our choice of weighted generalized mean and a simple mean, where the difference can be interpreted as an "unevenness penalty". The largest 10 such penalties are shown. While some companies have more than one zero, they all have zero signals for Percent Women on Board. Also 8 out of 10 of these companies are in the Oil, Gas & Coal Industry.

Company	BICS1	BICS2	G Signal	Board Tenure	Age Range	Chair Tenure	Pct Indep	Brd Avg Age	Pct Women	Average	Unevenness Penalty
Cognex	Industrials	Electrical Equipment	0.34	0.89	0.98	0.01	0.35	0.33	0	0.43	0.09
TransDigm Group	Industrials	Aerospace & Defense	0.66	0.91	0.98	0.74	0.91	0.94	0	0.75	0.09
Andeavor Logistics	Energy	Oil, Gas & Coal	0.41	0	0.96	1	0	1	0	0.49	0.08
Dril-Quip	Energy	OG&C	0.68	0.63	1	0.99	0.98	0.97	0	0.76	0.08
Western Midstream	Energy	OG&C	0.4	0	1	0.96	0	0.91	0	0.48	0.08
Diamond Offshore Drilling	Energy	OG&C	0.33	0.9	0.93	0.17	0.24	0.2	0	0.41	0.08
Continental Resources/OK	Energy	OG&C	0.47	0.91	0.4	0.66	0.99	0.32	0	0.55	0.08
Phillips 66	Energy	OG&C	0.41	0	0.98	0.93	0	0.99	0	0.48	0.07
PDC Energy	Energy	OG&C	0.58	0	1	1	0.92	0.99	0	0.65	0.07
Energy Transfer	Energy	OG&C	0.42	0.36	0.89	0.86	0	0.84	0	0.49	0.07

Table 19: Companies whose G signal differs the most from a simple mean for 2017. The "penalty" column is the difference between the simple mean of the 6 input signals and the generalized mean used in the G signal.

Interpretation of ES signal results is limited due to missing data. The 3 companies with the highest ES signal levels in each of the BICS 1 sectors is shown in Table 20. Notably, none of these have fatality rate data, and most of them do not have Spills or Safety data, so most of the "ES signal" is driven by the GHG Emissions and Energy use data in the Carbon Signal. Moreover, since we did not introduce any treatment or penalty for missing data, it is possible these companies benefit from missing data.

Company	BICS1	BICS2	Revenue	2018 Vol (demeaned)	ES Signal	Carbon Signal	Spills Signal	Fatality Signal	Safety Signal	CR&R Signal
Univar Solutions	Materials	Chemicals	8.3	11.9	0.99	0.99				
Valvoline	Materials	Chemicals	2.1	-16.4	0.98	0.98				
International Flavors & Fragra	Materials	Chemicals	3.4	-17.3	0.89	0.89				
Raytheon	Industrials	Aerospace & Defense	25.3	-2.1	0.80	0.80				
Apache	Energy	Oil, Gas & Coal	5.9	-8.5	0.79				0.60	1.00
Kinder Morgan/DE	Energy	Oil, Gas & Coal	13.7	-22.8	0.78		0.67			1.00
Schlumberger	Energy	Oil, Gas & Coal	30.4	-20.3	0.76	0.86				0.50
Itron	Industrials	Electrical Equipment	2.0	-8.2	0.73	0.81				0.50
Keysight Technologies	Industrials	Electrical Equipment	3.2	0.43	0.71	0.71				

Table 20: The top 3 ES signals in each BICS1 sector for 2017.

In order to better understand the whether or not a larger number of input signals are lowering signals, we show the average signal level among observations where there is sufficient data to construct 1, 2, 3 and 4 signals in Table 21. (Note no observation has the complete set of 5 signals.) On the whole, the answer seems to be no – though nominally the average ES signal decreases slightly as the number of signals increases, this is not statistically significant as the standard deviations for each bucket are in the range of 0.1-0.25.

Num Signals	**Num Obs**	**Carbon Signal**	**Spills Signal**	**Fatality Signal**	**Safety Signal**	**CR&R Signal**	**ES Signal**
0	625						
1	434	0.50	0.40	1.00	0.82	0.07	0.50
2	105	0.41	0.57	0.79	0.65	0.25	0.44
3	34	0.37	0.50	0.34	0.65	0.63	0.44
4	26	0.39	0.37	0.61	0.81	0.80	0.48

Table 21: Average signal values depending on available number of available ES signals per observation.

Size Bias in Signal Levels

It is widely accepted that larger companies have better ESG characteristics. This is perhaps because larger companies garner more attention from institutional investors who are actively monitoring governance practices and perhaps pursuing sustainable investing goals. As such we look for correlation between our signals and firm revenue. We find that the rank correlation between the G signal and revenue is +0.14 and significant. This is visualized in Figure 15.

However, regressing ES signals on revenue results in insignificant betas. Indeed we might expect size bias to be immaterial for ES because most of its univariate signals were constructed by normalizing on revenue.

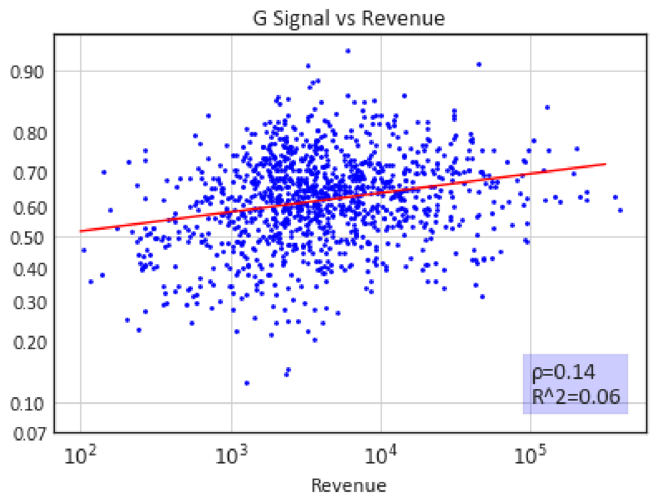

Figure 15: Relationship between revenue and G signal, showing positive correlation.

Size Bias in ES Disclosure

While we were able to eliminate size bias for ES via the signal construction, there is evidence that size bias remains for whether or not a company *discloses*. Table 22 shows that in all cases except Iron & Steel, the average revenue for companies disclosing enough to compute a signal is higher than companies that do not.

	Mean Revenue	
Industry	**No ES**	**Have ES**
Aerospace & Defense	14,845	27,579
Chemicals	4,149	9,174
Construction Materials	2,459	5,593
Containers & Packaging	6,281	7,061
Electrical Equipment	5,227	13,617
Industrial Services	4,578	7,092
Iron & Steel	**7,769**	**5,099**
Machinery	5,359	13,499
Metals & Mining	810	7,459
Oil, Gas & Coal	8,069	32,228
Transportation Equipment	5,418	16,242

Table 22: Mean revenues broken out by revenue and whether or not an ES signal could be constructed.

Is the G Signal Distinct from Size Bias?

Since larger firms tend to have less risk and better G signals as per our construction, it is reasonable to wonder if the two are statistically distinct. The evidence points to a qualified yes, based on comparing stand-alone and joint regressions as shown in Table 23. In the joint regression estimate, the G signal coefficient is significant at 5.9%. Recall that we did not adjust for size in the estimation of G signal slopes. One might expect that if we had explicitly fit the univariate G signal shapes on size-normalized volatilities the significance result would be sharper.

Regressors	**Slope Log Revenue**	**Slope G Signal**	**R^2**
Log Revenue	-8.6 [-10.1, -7.1]	-	0.097
G Signal	-	-14.5 [-21.1, -7.5]	0.016
Joint	-8.3 [-9.8,-6.7]	-6.3 [-12.8, 0.24]	0.100

Table 23: Regression tests for joint dependence of demeaned volatility on size and G signal. The p value for the slope of G Signal in the joint regression is 0.059.

Out of Sample Results

2018 data became available during the final preparation of this paper. This provided the opportunity for out-of-sample analysis. We used all the parameters (for both univariate and aggregate signals) as described above (and depending on data only through 2017), computed signals based on 2018 data, and compared the results to 2019 weekly volatility in Table 24. While there is not a monotonic decrease in volatility as G signals increase, it remains true that the top G quintile is less risky than the first, and that the average of the top two quintiles is lower than the average of the bottom two. Referring back to Table 15, 2016 has a similar pattern to this 2018 out-of-sample performance.

Note that since board characteristics do not change much year-over-year, neither do G signals – the standard deviation of the changes from 2017 to 2018 G signal differences is 0.079, and the R^2 of regression 2018 on 2017 is 0.985. Volatility is also somewhat serially correlated, as regressing the industry-demeaned 2019 volatilities on 2018 has an R^2 of 0.27. Given there is some stability of both quantities, we might have expected reasonable out-of-sample performance.

Year	G Q1	G Q2	G Q3	G Q4	G Q5
2018	-1.9	0.9	7.5	-1.3	-5.6

Table 24: 2019 volatilities (demeaned by industry) by 2018 G Signal quintiles.

Conclusions

In this paper we developed systematic approaches to extract signals from raw ESG data. Such signals are of interest for meeting sustainable investing goals while minding financial performance. The approach is general enough to incorporate sustainability goals into both the signal definition and aggregation processes. It is also amendable to validating design choices based on available historical data. This structured approach is of interest to investors and asset managers, especially since they are often tasked with implementing different objectives. ESG integration is not a one-size-fits-all process.

Signal design is guided by a framework that associates individual fields to ESG factors. These frameworks embody sustainable development goals and acknowledge differences in financial materiality among industries. Multiple frameworks are another manifestation of the lack of consensus on objectives, and suggests this layered signal approach can be used to implement different mandates.

Indeed, our approach consists of constructing a univariate signal from single ESG fields, including accounting for economic scale in fields such as emissions, spills and fatalities. One conclusion from our analysis is that many fields exhibit a decreasing marginal impact – larger firms enjoy a smaller per unit impact than smaller firms – implying that intensity ratios can be biased. For G fields, evidence shows that financial performance is improved using non-monotonic signal curves. By leveraging the length of the G data, we determine "sweet spots" where the signal curve peak, at least based on historical performance.

Once univariate signals are constructed, factor analysis is used to determine if different signals are measuring the same factor. Factor results are compared to the framework to ensure that signals combine usefully. When univariate signals behave as a single factor, we use simple averages as to cancel measurement error as well as to serve a proxy for missing input signals. Once such signals are combined, the remaining factors are truly multivariate. Here we employ generalized means to implement the notion that good performance in one factor should not fully compensate poor performance in another. As with G signal design, we estimate aggregation parameters using historical data, establishing statistical evidence for this "non-compensability". However, we also showed that it one can set the generalized mean parameter to reflect one's views via stylized examples. In either case, the approach provides incentives for even performance in a manner that weighted averages cannot.

Analysis shows that aggregated ESG signals are negatively correlated with return volatility. This paper therefore provides further evidence – directly from data, not ratings – that better ESG performance can result in better portfolios. Importantly, the approach provides important transparency into what drives aggregate signals, and other methods of aggregation could correlate with return volatility differently while perhaps better fulfilling some other purpose.

By highlighting the subjective design choices made here, we hope some light has been shed on reasons why there is divergence among ESG score providers today. As demand for ESG integration increases, and objectives become clearer, it is necessary to have transparency about the methodology, especially as to how bias is treated and how multi-factor ESG signals are designed to match objectives.

These are early days for ESG integration. For ES signals in particular, reporting is low and disclosure itself has a size bias. Nonetheless it is our hope that by offering a structured, data-driven signal methodology, more companies will be encouraged to disclose, making it easier for investors to construct portfolios that truly reflect their objectives.

Acknowledgements

The author would like to thank members of the Bloomberg Sustainable Finance Solutions, Global Data, BI ESG, Quant Research and Index Research teams for numerous conversations and valuable insights leading to the development of ideas presented here. Scope and details of the case study, along with any remaining errors, remain the responsibility of the author.

References

Berg, F., J. Kölbel and R. Rigobon. "Aggregate Confusion: The Divergence of ESG Ratings." *MIT Sloan Research Paper* (2019).

Bloomberg LP. "2020 Bloomberg Gender-Equality Index Methodology ." 2020.

Ellis, G. and A. Fisher. "Valuing the Environment as Input." *Division of Agricultural Sciences, University of California Working Paper* (1986).

Fullerton, D. and Chi. Ta. "Environmental Policy on the back of an envelope: A Cobb-Douglas model is not just a teaching tool." *Energy Economics* (2019).

Henriksson, R., et al. "Integrating ESG in Portfolio Construction." *Journal of Portfolio Management* (2019): 67-81.

Huang, S. and G. Hilary. "Zombie Board: Board Tenure and Firm Performance." *Journal of Accounting Research* (2018).

Khan, M., G. Serafeim and A. Yoon. "Corporate Sustainability: First Evidence on Materiality." *Accounting Review* (2016).

Kotsantonis, S. and G. Serafeim. "Four Things No One Will Tell You About ESG Data." *Journal of Applied Corporate Finance* (2019): 50-58.

Kotsantonis, S., C. Pinney and G. Serafeim. "ESG Integration in Investment Management: Myths and Realities." *Journal of Applied Corporate Finance* (2016): 10-16.

Meyers. "Predictive Modeling with the Tweedie Distribution." *CAS Annual Meeting*. 2009.

OECD. "Handbook on Constructing Composite Indicators: Methodology and User Guide." 2008.

Rahdari, A. and A. Rostamy. "Designing a general set of sustainability indicators at the corporate leve." *Journal of Cleaner Production* (2015).

Yang, Y. "Insurance Premium Prediction." *arxiv.org* (2016).

York, R., E. Rosa and T. Dietz. "STIRPAT, IPAT and ImPACT: analytic tools for unpacking the driving forces of environmental impacts." *Ecological Economics.* (2003).

Appendix

In order to unify the next two sections with Emissions modeling, we first point out that in all cases the goal is to estimate the parameters C and γ in the relationship

$$\mathbf{E}[\text{Impact}] = C \times \text{Activity}^{\gamma}$$

where $C = \exp[\text{Average}\log\text{Intensity}]$. In sum, the treatment depends on the distribution of the dependent variable, Impact. The three formulations (Lognormal, Negative Binomial and Tweedie) are all flavors of *exponential dispersion models*, which unify ordinary least squares regression with these families (and a few more).

We also mention that the problem of modeling skewed count and amount data naturally occurs in the insurance industry, where the number of claims is the count variable, and the losses are the amount variable. Moreover, actuaries have a keen need to determine not only the expected claim counts and losses as a function of predictor variables, but to understand the distributions as well – see, for example, (Meyers) and (Yang). For certain types of environmental risks, it seems natural to adopt similar techniques for predicting the probabilities of different outcomes.

Count and Loss Data

Let's first turn our attention to case of Hazardous Materials Spills, for which we have both count of spills and spill amounts[8]. First of all, note that for both counts and amounts, zero is the *desired* level and is not in the middle of the distribution. Moreover, even if zero is not the most common outcome (i.e. the mode), we might hope to find a cluster of small values. In such cases, it will not be possible to transform such data to approximate normality. This artefact of the data is shown in Figure 16 for spill counts where 12% of the observation have spill counts of 0 or 1 and the Kolmogorov-Smirnov test for normality rejects normality with a p-value of 0.004.

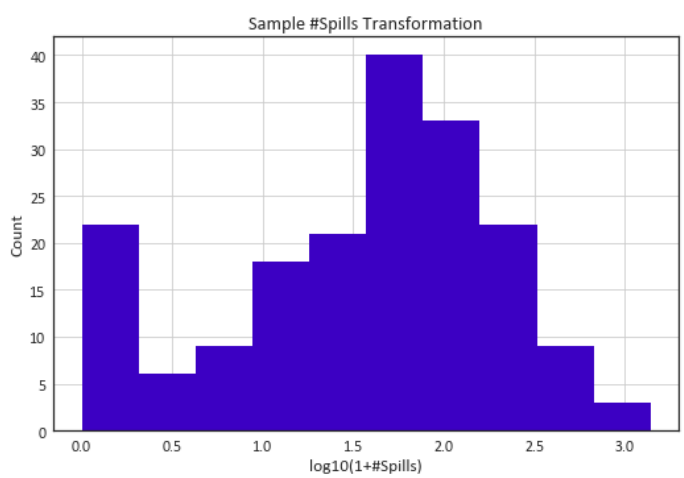

Figure 16: Histogram of #Spills, following a Box-Cox transformation of log(1+x). Note the cluster of small values is unlikely to go away under any Box Cox transformation, meaning that one cannot use a transformation to approximate normality.

Consider again the formulation

$$\text{Impact} = \text{Intensity} \times \text{Activity}^{\gamma}$$

Since the dependent variable is integer-valued, instead of assuming log Impact is a

[8] In practice we might also have recovered amounts. In that case, one might subtract the recovered amounts from the spill amounts, making the spill count play the role of differentiating between zero spills and perfect clean-ups. We omit this nuance in our case study.

normal distribution with $\mu = \text{Average log Intensity} + \gamma \times \log \text{Activity}$ and $\text{Var}(\varepsilon) = \sigma^2$, we assume Impact follow a negative binomial distribution with

$$r = \frac{\mu^2}{\sigma^2 - \mu}$$
$$p = \frac{\mu}{\sigma^2}$$

As before, μ depends on log Activity, γ and a constant capturing an average intensity. Moreover the signal is calculated according to the corresponding distribution function

$$1 - \text{NegBin}(y; r, p)$$

However, unlike before, the parameters now need to be estimated with more sophisticated techniques such as maximum likelihood.

Results of this estimation process are shown in Table 25 and Figure 17.Not only do the data clearly exhibit diminishing marginal impacts, but the 25% percentile remains very low, indicating that good performance requires small spill counts even for large companies. It is also remarkable that the very largest spill counts occur for some of the lower-revenue companies.

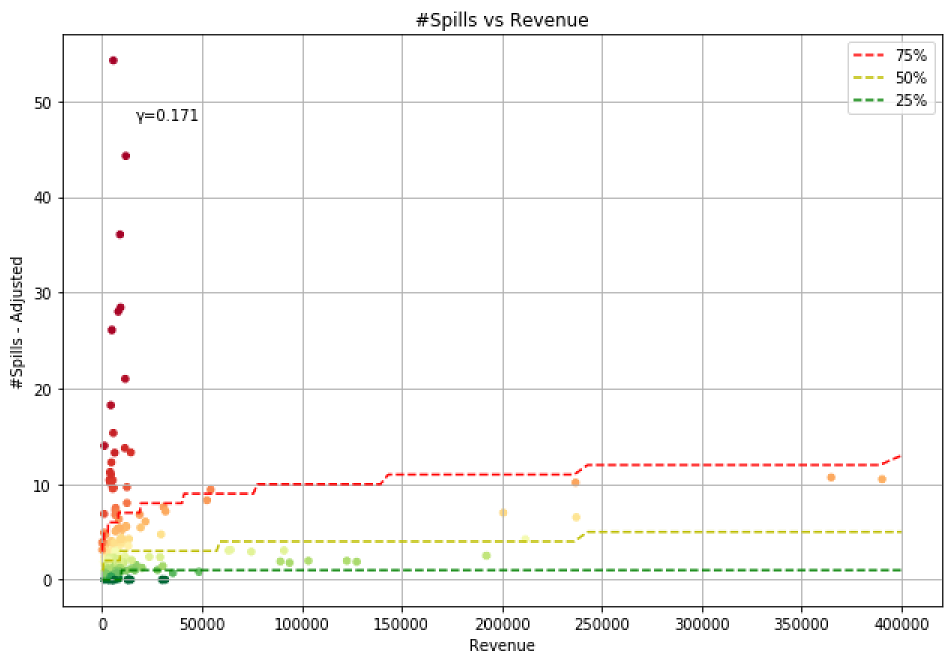

Figure 17: Signal extraction for number of spills adjusted by sector-specific unit intensities. As before green dots indicate signal levels near 1 and red rots indicate signals near 0.The estimated elasticity of γ=0.171 indicates strongly diminishing marginal impacts.

Industry	Intercept	Unit Intensity
Energy	3.45	31.4
Materials	2.41	11.2
Industrials	0.26	1.3

Table 25: Results of Negative Binomial Regression for #Spills versus Revenue, where the estimated elasticity is γ=0.166 and μ=exp(intercept+ γ log(Revenue)).

Of course, the number of spills is only part of the story – the *amount* of the spills also matters. Spill amounts share similar statistical characteristics with counts, except that the dependent variable is now a real number and not an integer. One appropriate treatment for such data is the Tweedie distribution, which for certain parameter choices can be thought of as the distribution generated by summing up a random number of non-negative random variables.

$$Y = \sum_{j=1}^{N} A_j$$

Here N is Poisson(λ) and each A_j is Gamma(α,θ). The standard Tweedie parameters are

$$p = \frac{\alpha + 2}{\alpha + 1}$$
$$\mu = \lambda\alpha\theta$$
$$\phi = \frac{\lambda^{1-p}(\alpha\theta)^{2-p}}{2 - p}$$

For this compound Poisson-Gamma distribution, one has $1<p<2$ and $\mathrm{Var}[Y] = \phi\mu^p$. It remains the case that μ depends on log Activity, γ and a constant capturing an average intensity.

Results of a Tweedie fit for our sample data is shown in Figure 18. We see that both spill counts and amounts exhibit diminishing marginal impacts, meaning that their spills do not increase linearly with revenue. Had we simply used intensities, the resulting signals for large companies would have been biased upwards.

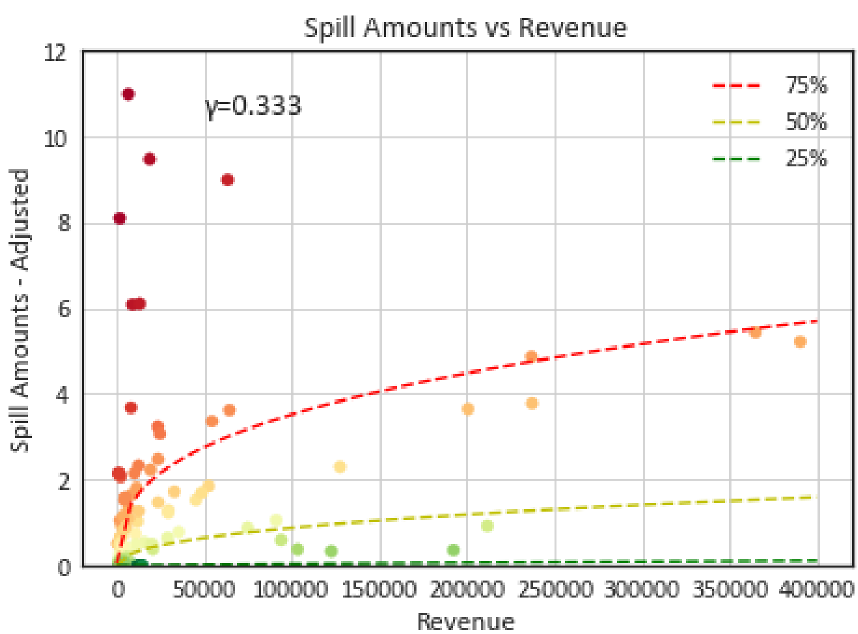

Figure 18: Signal extraction for spill amounts. Again we see evidence for diminishing marginal impacts as companies scale. Units are thousands of metric tons. Note: roughly 10% of the data has values greater than 12 (and are not shown). These were also excluded from estimation, as the MLE method used is not particularly robust to outliers.

Fatalities

In the case of fatality rates, we do have the numerator and denominator (fatalities and number of workers), reported separately for employees and contractors. In light of the discussion concerning spill incidents, we recognize that the data is likely to have a large proportion of zeros, and that the fatality rate may not have unit elasticity. To that end we apply a Zero-Inflated Poisson model, in which the event rate can depend on the number of workers. Additionally, this model has an additional probability mass at zero, controlled by θ. With the dependent variable Y being the number of fatalities, the model is

$$\mu = \exp(c + \gamma \log(\#\text{Workers}))$$
$$P(Y = y) = \begin{cases} \theta + (1-\theta)\text{Poisson}(0|\mu) & y = 0 \\ (1-\theta)\text{Poisson}(y|\mu) & y > 0 \end{cases}$$

Similarly to Negative Binomial and Tweedie models, Zero-Inflated Poisson models can be estimated via maximum likelihood. As with the incident rates, we pool together observations from all industries in our sample to reflect the principle that human capital has the same value in all industries.

Results of the fatality estimation are show in Figure 19. Indeed we see that the elasticity is far less than one: the expected number of fatalities does not increase at the same rate for large and small companies.

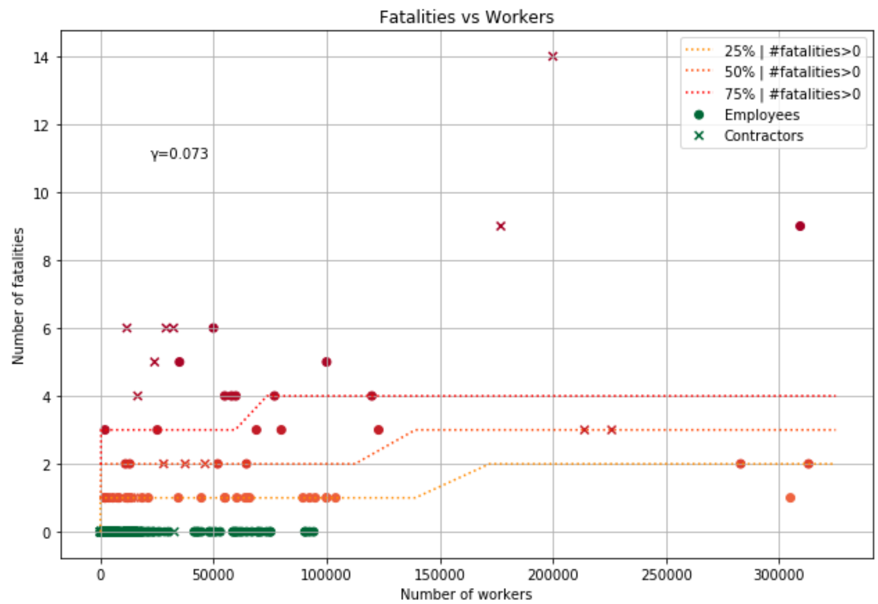

Figure 19: Fatalities versus size of workforce. Both employees and contractor fatalities are shown. Fatality rates clearly do not scale with unit elasticity.

Important Disclosures and Disclaimer:

Any systematic investment strategies described herein may involve a high degree of risk, including without limitation market risk and other risks inherent in investing in securities, commodities, currencies, derivatives and other financial instruments. The value of and income from investments linked to such strategies may decline in value and loss of the original amount invested can occur. All levels, prices and spreads are historical and do not represent current market levels, prices or spreads, some or all of which may have changed since the publication of this document.

Bloomberg does not represent that the index data, quantitative models, analytic tools and other information ("Content") referenced in this publication (including information obtained from third party sources) is accurate, complete or error free, and it should not be relied upon as such, nor does Bloomberg guarantee the timeliness, reliability, performance, continued availability, or currency of any Content. The Content is provided for informational purposes only and is made available "as is." Because of the possibility of human and mechanical errors as well as other factors, Bloomberg accepts no responsibility or liability for any errors or omissions in the Content (including but not limited to the calculation or performance of any index and/or the output of any quantitative model or analytic tool). Any data on past performance, modelling or back-testing contained in the Content is no indication as to future performance. No representation is made as to the reasonableness of the assumptions made within or the accuracy or completeness of any modelling or back-testing.

Bloomberg shall not be liable for any damages, including without limitation, any special, punitive, indirect, incidental or consequential damages, or any lost profits, arising from the use of or reliance on any Content, even if advised of the possibility of such damages.

Indices are unmanaged and cannot be invested in directly. The development or creation of any product that uses, is based on, or is developed in connection with any index (each a "Product") is prohibited without the prior written consent of Bloomberg. Bloomberg does not sponsor, endorse, sell or promote such Products and makes no representation regarding the advisability of investing in any such Product. Index returns represent past performance and are not indicative of any specific investment. The Content (including any of the output derived from any analytic tools or models) is not intended to predict actual results, which may differ substantially from those reflected.

Information and publications provided by Bloomberg shall not constitute, nor be construed as, investment advice or investment recommendations (i.e., recommendations as to whether or not to "buy", "sell", "hold", or to enter or not to enter into any other transaction involving any specific interest) or a recommendation as to an investment or other strategy. No aspect of the Bloomberg publications is based on the consideration of a customer's individual circumstances. Information provided in the publications should not be considered as information sufficient upon which to base an investment decision. You should determine on your own whether you agree with the conclusions made in the publications.

Take the next step.

For additional information, press the <HELP> key twice on the Bloomberg Terminal®.

bloomberg.com/professional

Beijing
+86 10 6649 7500

Dubai
+971 4 364 1000

Frankfurt
+49 69 9204 1210

Hong Kong
+852 2977 6000

London
+44 20 7330 7500

Mumbai
+91 22 6120 3600

New York
+1 212 318 2000

San Francisco
+1 415 912 2960

São Paulo
+55 11 2395 9000

Singapore
+65 6212 1000

Sydney
+61 2 9777 8600

Tokyo
+81 3 4565 8900

Bloomberg offers its services in compliance with applicable laws and regulations. Services and information provided by Bloomberg should not be construed as tax or accounting advice or as a service designed to facilitate any subscriber's compliance with its tax, accounting, or other legal obligations. Employees involved in Bloomberg may hold positions in the securities analyzed or discussed in Bloomberg publications. This publication has been produced and distributed by Bloomberg. "Bloomberg" means Bloomberg Finance L.P., Bloomberg L.P., Bloomberg Index Services Limited ("BISL") and their affiliates. BISL is an independent wholly owned subsidiary of Bloomberg L.P. As the administrator of the Bloomberg Barclays family of indices, BISL operates independently from Bloomberg L.P. And its other affiliated entities.The data included in these materials are for illustrative purposes only. ©2020 Bloomberg L.P. S808215017 0817