Using Alternative Data to Improve ESG Insights

Much has been written about the challenges of ESG data. While well-reasoned in their construction, the ratings produced by the large ESG data vendors are not as comparable as one might hope, leading to skepticism and a debate over which companies are actually 'good' from an environmental, social, or governance perspective. Skipping these headline ratings in favor of the data that underlies them is a natural reaction (and is the practice of most sophisticated investors) but focusing on the raw data does not solve three key problems: paucity, 'believability', and ubiquity.

Given the challenges of traditional ESG data, investment practitioners have a strong incentive to look to alternative data sets for help. These data step away from what is reported by companies and captured in the traditional, structured ESG data sets produced by MSCI and the like, and instead focus big data and/or unstructured data. This is the data we find on websites, social media, physical sensor data, satellite data, images, and videos, for example. By most estimates, 80-90% of data produced on companies fall into the unstructured data category, therefore we have an enormous incentive to dig for information here! Importantly, though, the usefulness of these data depends on our ability to extract actionable information from them --for this we turn to a host of quantitative methods from machine learning to natural language processing (NLP), and even blockchain.

We have found that these new, alternative data allow us to start to chip away at the three problems mentioned above. Specifically, we can use these new sources of information for augmentation of traditional datasets, addressing the paucity problem. Within E, S, and G there are lots of omissions in company-reported, structured data (such as one would find in CSR reporting, or even in some regulatory reports).   Because most ESG reporting is voluntary, most companies reveal only a subset of items, and may do so in non-standardized ways. Data on workplace safety and protections, diversity, and other aspects of the relationship between labor and management are examples of 'S' data that are often underreported (and therefore missing) for many companies. We could capture this 'social' information through web scraping or from sites like Glass Door, Twitter, or even government and NGO websites to help us paint a more complete picture of what's actually going on inside companies and fill in some of the known 'holes' in traditional data.

Validation of ESG information is critical to our practice. The 'believability' of company claims is often in question, especially in the absence of multiple data sources. For example, companies have adopted myriad policies to combat destructive environmental practices (e.g. threats to biodiversity), ensure data privacy, or prohibit questionable labor practices. As an outside observer of the company, it is difficult to gauge the effectiveness of these policies (or even to what extent management takes them seriously). We can look to alternative data to help us gain

confidence in the typically structured, company-reported information that forms the backbone of much of our work. Company supply chains are notoriously opaque, for example, and companies claiming to have supply chain practices that are free from child- or slave-labor, or even free from commodities sourced in precarious ecosystems are difficult to validate –both for investors and often the companies themselves. We are excited about various efforts underway to use blockchain to validate supply chain claims. With the origins and movements of goods recorded immutably in a distributed ledger we could have far greater confidence in programs aimed at mitigating supply chain risk. Initiatives like these are already at work in the precious metals and gemstones, minerals mining, automotive, food (food safety), and apparel sectors.

The ubiquitous nature of company-reported, structured data is problematic. By virtue of the fact that it forms the basis for most of the large ESG data vendors' ratings, it means that all investors are, de facto, using the same data. As active managers, we seek novelty when it comes to information. ESG is no different. Our goal is to produce the most robust view of the threats and opportunities faced by companies – new or proprietary data sources can give us an edge when it comes to better identifying ESG-related down-side risks or upside potential. Further, novel data – especially that which is not under the control of companies – is a powerful source of diversification within our data pool. ESG news flow analysis naturally falls into this category – not only by virtue of being a new data source, but more importantly being a new concept, when compared to traditional ESG data. By using natural language processing (NLP) paired with deep subject matter knowledge, we can produce what is essentially 'ESG sentiment' information that is based on a wide number of text-based data sources. Again, the beauty of this information is that it represents a genuinely new concept and is largely independent of what companies say about themselves.

But alternative data itself is not without its challenges. First and foremost, the bulk of this data has not been road tested in an investment context. We must approach the data by asking what fundamental economic question it answers, and be circumspect about what it will bring to our process. Next, it is important to acknowledge that this data can be very expensive. Data vendors are pouring into this space because 1) the technical barriers to entry are much lower than they were even five years ago, and 2) they know that managers will pay big bucks for data sets that might give them an advantage (even a reputational one). Finally, we see many examples of investors being 'blinded by technique'. While there is truly astonishing work going on in the field of Artificial Intelligence (AI), for example, it doesn't mean that the work will necessarily result in information that will add to our understanding of companies' ESG attributes.

In sum, we are excited and cautiously optimistic about what alternative data can add to our understanding of companies' ESG practices. We would never suggest using alternative ESG data in isolation, but combined with traditional, structured data sets we believe it could result in a much more robust understanding of the true threats and opportunities that companies face.

Kathryn McDonald, Co-Founder and Head of Investments and Sustainability, Radiant Global Investors