The Rise of AI in ESG Evaluation

July 2019

Citi Global Data Insights
Citi Global Data Insights Team (cgdi@citi.com)
# Table of Contents

The Focus on ESG .............................................. 4

Motivation and Scope ....................................... 7

AI in ESG .................................................. 9

   RepRisk ............................................... 10

   Truvalue Labs ......................................... 26

   Arabesque S-Ray ...................................... 41

Analyst-Driven vs AI-Led .................................. 59

Conclusion .................................................. 63
The Focus on ESG

The concept of ESG (Environmental, Social and Governance) or Sustainable Investing has been around for decades, tracing back to the 1950s and 60s where trades unions recognised their invested capital in companies could have wider implications socially. However, the focus on ESG has been markedly rising in the investment industry only in the past 5 years or so.

There is clear evidence of the recent upward trend based on the number of signatories declaring their adherence to the United Nations’ six Principles for Responsible Investment (PRI). Specifically the signatories increased significantly since 2014, doubling from 1200 to almost 2400 institutions, including asset owners (AO) and asset managers. The total AUM as depicted below also jumped from US$ 45 trillion to close to US$ 90 trillion for the same period.

![Figure 1. UNPRI Signatories and AUMs](source: UNPRI, CGDI)

This shift is likely to have been driven by strong client demand as investors pay more attention to ESG related issues, e.g. climate change, coupled with research evidence showing that complying with ESG does not detract from performance and in some cases actually add performance. A recent event study by Grewal, Ridel and Serafeim (2017) focused on the effect of implementation of EU directive 2014/95 on companies as mandatory reporting in ESG issues came into force. Their results show that firms’ ESG performance has direct linkage to market reactions where more positive returns are observed for firms with better ESG performance and vice versa, especially in environmental and governance areas.

On the demand side, the search volume of the term "ESG investing" has witnessed an exponential increase pattern since 2015¹. The surging interest in ESG investing could potentially be attributed to the following important ESG developments: (1) the launch of the UN’s 17 Sustainable Development Goals (SDGs), (2) the discussion, signing and adoption of the Paris Climate Agreement by 196 countries, (3) prominent asset owners becoming signatories of the UN’s Principles for Responsible Investment (PRI), e.g. Japan’s Government Pension Investment Fund.

¹ Based on Google Trends
Arguably investors’ attention has accelerated the growth of ESG-focused funds. Indeed, through our conversations with investors, it is clear that the importance of embracing ESG in their investment processes cannot be overstated. In fact, the EMEA Head of Active Equities from one of the largest asset managers in the world stated that their ESG efforts deployed in their funds are now the prerequisite for any meetings.

“In all recent meetings we have had with pension funds, every one of them wants to know how we incorporate ESG in the funds. Without it, they would not even take a meeting”

Two large US state pension funds seemingly echo the requirement of ESG implementation, as they stated that all of their external managers are asked to incorporate ESG as one of the key risk factors. Interestingly they also state that amongst their managers, some have started to include positive screening for ESG leaders.

Based on the institutional fund database provided by eVestment, the percentage of “ESG” searches relative to the total searches by asset consultants (see Figure 3, measured as 3-month moving average) has also increased nearly six-fold since 2017. In addition, Figure 4 overleaf demonstrates that the number of ESG-focused funds (including both equity and fixed income) has more than doubled, and their total AUM had reached nearly US$ 500 billion at the end of 2018.
With the increasing interests in ESG investing, more and more companies are making selective and unaudited disclosures in order to attract ESG-investing capital. Consequently, ESG rating or scores providers play a crucial role in assessing the impartiality of such disclosures to help investors navigate through a myriad of information.

However, individual providers’ ESG ratings/scores can vary significantly, due to differences in methodology, subjectivity in interpreting information from disclosures, and their own research scope. Such variations lead to inconsistent applications and understanding of ESG ratings by asset managers when they incorporate them into their portfolios. While this does not mean there is any one particular way in ESG investing that is right or wrong, it shows that such investment decisions often remain more art than science.
Motivation and Scope

In order to help asset managers implement a more structured way to incorporate ESG we need to understand how ESG is currently being used. Based on the widely accepted global standard of ESG/Sustainable Investing classification, the definitions published by Global Sustainable Investment Alliance (GSIA) in 2012 consist of the following:

- **Negative/Exclusionary Screening**
  Exclusion of certain sectors, companies or practices based on specific ESG criteria

- **Positive/Best-in-class Screening**
  Investment in sectors, companies or projects based on positive ESG performance relative to industry peers

- **Norms-based screening**
  Screening against minimum standards of business practice based on international norms

- **ESG Integration**
  Systematic and explicit inclusion by asset managers of ESG factors into financial analysis

- **Sustainability Themed Investing**
  Investment in sustainability related themes specifically

- **Impact/Community Investing**
  Investment aimed at solving social or environmental problems, including community investing

- **Corporate Engagement and Shareholder Action**
  The use of shareholder power to influence corporate behavior guided by comprehensive ESG guidelines

![Figure 5. Sustainable Investing Assets by Strategy and Region 2018](source)

Traditionally ESG screening has focused more on negative screening; excluding companies or industries that are perceived to be contradictory to ESG principles (e.g. the tobacco industry). Figure 5 and Figure 6 from GSIA show that while negative screening is still the most common practice in ESG investing, especially in Europe, "ESG Integration", "Positive/Best-in-class Screening" and "Sustainable themed investing categories" have all experienced significant asset growth.

In order to perform ESG analysis, regardless of which definition of sustainable investing asset managers follow, the key challenge is finding sources of data they can use to help evaluate companies' behavior along the ESG spectrum within their investable universe.
The main source of information is the ESG disclosure made by companies which typically occurs annually. As asset managers often do not have large in-house analyst teams focusing on ESG, the majority of them rely on ESG index providers who analyze company disclosed information and evaluate firms using methods similar to those in financial modelling.

While this method appears reasonable, it suffers from several drawbacks:

- Lack of consistency and comparability in companies’ disclosed information
- No common reporting standards mandating companies to disclose
- Difficulty in measuring companies’ intra-year ESG performance to take actions based on annual disclosures
- Human analysis making assessments less timely and volume limiting
- Lack of transparency in how ESG index providers derive their scores

These issues have undoubtedly accelerated the rise of AI-based ESG providers we have witnessed in recent years as investors consider alternative ways to address these issues.

In the era of data abundance, these AI-led ESG providers ingest unstructured data and texts through natural language processing (NLP) and machine learning to provide an up-to-date view on a company’s ESG performance from an “outside-in” perspective. Companies are assessed based on what the outside world says about their business activities along environmental, social and governance dimensions. Consequently, asset managers are no longer solely reliant on company-disclosed information about their ESG behaviors. And since there are no binding standards or mandatory line-item reporting requirements related to ESG information, companies can choose to what to disclose and what not to disclose. For example, Huffington Post (Dec 2017) reported that 90% of negative ESG events were not disclosed in either the SEC filings or sustainability reports. In the case of AI-led ESG providers, such free optionality would have no bearing on their scoring of companies. Automated scoring can lessen the potential for biases and enable analysis at scale.

However, AI-led ESG providers have a different challenge: finding meaningful, material and predictive information out of enormous amount of unstructured data. It goes without saying that this challenge introduces issues surrounding data structure, taxonomy, identifier mappings, modelling accuracy, coverage, comparability, and data delivery etc.

In this report, we highlight three different AI-led ESG providers where one focusses on the downside risk, another one on scoring companies from both positive and negative sides with varying investment horizons taken into consideration, and the last one aggregates data from other third party providers to derive ESG scores. The objective of the report is to discuss each vendor's methodology, construction techniques and how their AI / machine learning approaches address the aforementioned challenge. We will also perform an initial empirical analysis of their scores to understand the tilts, coverage and examine performance. The main investable universes we use for this analysis are MSCI World and MSCI Emerging Market indices.

---

**Figure 6. Global Growth of Sustainable Investing Strategies 2016 – 2018**

<table>
<thead>
<tr>
<th>Year</th>
<th>Impact/community investing</th>
<th>Sustainability investing</th>
<th>Positive/best-in-class screening</th>
<th>Norms-based screening</th>
<th>Corporate engagement and shareholder action</th>
<th>ESG Integration</th>
<th>Negative/exclusionary screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>$444,26</td>
<td>$1,017,66</td>
<td>$1,841,37</td>
<td>$4,679,44</td>
<td>$9,834,39</td>
<td>$17,843,81</td>
<td>$19,770,95</td>
</tr>
<tr>
<td>2016</td>
<td>$248,47</td>
<td>$275,16</td>
<td>$318,01</td>
<td>$6,195,40</td>
<td>$6,385,17</td>
<td>$10,263,20</td>
<td>$15,063,57</td>
</tr>
<tr>
<td>Growth 2016-18</td>
<td>79%</td>
<td>269%</td>
<td>125%</td>
<td>-24%</td>
<td>-17%</td>
<td>-69%</td>
<td>-31%</td>
</tr>
<tr>
<td>CAGR</td>
<td>33.7%</td>
<td>90.0%</td>
<td>60.1%</td>
<td>-13.1%</td>
<td>8.3%</td>
<td>30.2%</td>
<td>14.8%</td>
</tr>
</tbody>
</table>

*Note: Asset values are expressed in billions.*

Source: Global Sustainable Investment Alliance, GSIR Review March 2018
AI in ESG

The three AI-led ESG vendors\(^2\) which we will explore in the following sections are RepRisk, Truvalue Labs and Arabesque S-Ray. They are chosen as their methodologies are distinctly different from each other. But they do have something in common: AI / Machine Learning techniques are deployed on large volumes of unstructured data is deployed to provide an outside-in perspective of companies based on ESG measures.

RepRisk has positioned itself as a specialist in monitoring and alerting material ESG risk incidents and violations of ten principles of the UN Global Compact (UNGC) of companies. They use NLP on 80,000 data sources to identify relevant ESG incidents based on their own defined 28 issues and 57 topics and themes, which are mapped to UNGC. Their core offerings are the RepRisk Index (RRI) and the RepRisk Rating (RRR).

TRUVALUE LABS

Truvalue Labs, on the other hand, assesses companies from both positive and negative perspectives. Through AI and big data analytics, they deliver objective algorithmic scoring on ESG factors as identified by the Sustainability Accounting Standards Board™ (SASB™) as having material impacts on company value by industry and by sector. The core offerings are Pulse, Insight, Momentum and Volume scores.

Arabesque S-Ray sources ESG data from a wide range of third party data vendors, categorised as "Report-based", "News-based", "NGO-based" and "Materiality". They then transform these four types of data through modelling and machine learning into scores. The core offerings are Global Compact (GC) score, ESG score and individual E/S/G Scores.

In the next sections, we will examine each provider in turn and discuss their methodology and construction of their ESG scores, followed by empirical analysis\(^3\) on their offerings.

---

\(^2\) Vendors listed here are either in talks or have been working with Citi.

\(^3\) Quintile portfolios are sector neutral, rebalanced monthly.
The Methodology and Construction sections have benefitted from discussions with RepRisk, with thanks to Danilo Chiono, Head of Sales, danilo.chiono@reprisk.com, Gina Walser, gina.walser@reprisk.com and team.

Headquartered in Zurich, Switzerland, and founded as Ecofact in 1998, RepRisk was originated in the credit risk department within a global bank. Since 2006, the company has been leveraging artificial intelligence and curated human analysis to translate big data into actionable business intelligence and risk metrics. At the time of writing, RepRisk has 120 employees globally, with 70 analysts who provide the human element in their research process.

With daily updated data synthesized in 20 languages using a rule-based methodology, RepRisk systematically flags and monitors material ESG risks and violations of international standards that can have reputational, compliance, and financial impact on a company. The RepRisk Platform covers 120,000+ public and private companies and 30,000 projects covering every sector and market. According to RepRisk, their clients utilise them as the main due diligence solution to assess ESG and business conduct risks related to their operations, business relationships, and investments.

---

**Figure 8. RepRisk ESG Score Process**

Artificial Intelligence
- Aggregator
- Big data: 500k news screened and pre-processed per day

Human Intelligence
- Admin
- Risk incidents are curated, analyzed, and given a severity score and summary

Due Diligence Tool
- RepRisk Platform
- Curated risk research, metrics, and analytics

>>> Daily updates >>>

Source: RepRisk
METHODOLOGY

Philosophy

RepRisk believes it is important to look at companies' ESG performance, not just their policies. By "performance", they mean whether or not companies do what they say on the ground by adopting an "outside-in" approach. To assess a company based on this approach, their research captures and analyses information from media, stakeholders, and other public sources external to the said company. This external perspective helps to gauge whether a company’s policies and processes are translating into real actions on the ground. In essence, RepRisk acts as a "reality check" of a company’s business conduct along the ESG dimension.

Sources of Data and Update Frequency

RepRisk screens over 80,000 public sources and stakeholders in 20 languages\(^4\) on a daily basis. These sources consist of print media, online media (including social media such as Twitter and blogs), government bodies, regulators, NGOs, think tanks, newsletters, and other online channels. The sources range from the international, regional, national to local level. The data history goes back to January 2007. This list of sources is reviewed regularly and extended according to daily searches, RepRisk’s own research, and through client feedback.

One thing to note is that RepRisk do not verify or validate the allegations captured from their sources; they believe their role is to provide relevant information transparently, and therefore the focus is on identifying and assessing the risk incidents in a systematic and rule-based way.

How AI / Machine Learning is Applied

RepRisk employs natural language processing (NLP) to identify relevant ESG issues/incidents according to their defined research scope (detailed in the Construction section later) and then map those incidents to the relevant company. In other words, their methodology is issue / event-driven, rather than company-driven. They then analyze risk incidents, add curated research and analytics to the platform, and update proprietary risk metrics whenever new risk information is published.

This means that RepRisk can potentially cover any company exposed to ESG risks captured by their data sources, regardless of the company’s size, sector, country of origin, or whether the company is public or private. RepRisk states that their coverage grows daily as new relevant information is identified and analyzed. Roughly 30-50 new companies are added each day. As of March 2019, the RepRisk Platform covers more than 120,000 companies that are associated with risk incidents. Of these companies, approximately 15% are listed firms and 85% are non-listed.

\(^4\) English, Arabic, Chinese, Danish, Dutch, Filipino, Finnish, French, German, Hindi, Italian, Indonesian (Bahasa Indonesian), Japanese, Korean, Malaysian (Bahasa Malaysian), Norwegian, Portuguese, Russian, Spanish, and Swedish.
CONSTRUCTION

ESG Criteria Used

RepRisk’s research scope is comprised of 28 ESG issues defined by them as broad, comprehensive and mutually-exclusive. The 28 Issues in Figure 9 below drive their entire research process, where every risk incident on RepRisk’s ESG Risk Platform is linked to at least one of these Issues. When RepRisk screens the sources and stakeholders, it screens for any company or project linked to these Issues.

Figure 9. RepRisk ESG 28 Issues

<table>
<thead>
<tr>
<th>ENVIRONMENT</th>
<th>SOCIAL</th>
<th>GOVERNANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental Footprint</td>
<td>Community Relations</td>
<td>Corporate Governance</td>
</tr>
<tr>
<td>Climate change, GHG emissions, and global pollution</td>
<td>Human rights abuses and corporate complicity</td>
<td>Corruption, bribery, extortion, money laundering</td>
</tr>
<tr>
<td>Local pollution</td>
<td>Impacts on communities</td>
<td>Executive compensation issues</td>
</tr>
<tr>
<td>Impacts on landscapes, ecosystems, and biodiversity</td>
<td>Forced labor</td>
<td>Misleading communication</td>
</tr>
<tr>
<td>Overuse and wasting of resources</td>
<td>Child labor</td>
<td>Fraud</td>
</tr>
<tr>
<td>Waste issues</td>
<td>Local participation issues</td>
<td>Tax evasion</td>
</tr>
<tr>
<td>Animal mistreatment</td>
<td>Social discrimination</td>
<td>Tax optimization</td>
</tr>
</tbody>
</table>

Cross-cutting Issues

- Controversial products and services
- Products (health and environmental issues)
- Supply chain issues
- Violation of national legislation
- Violation of International standards

Note: The 28 RepRisk ESG Issues map the Ten principles of the UN Global Compact

The issues are selected and defined in accordance with the key international standards related to ESG issues and business conduct, such as the World Bank Group Environmental, Health, and Safety Guidelines, the IFC Performance Standards, the Equator Principles, the OECD Guidelines for Multinational Enterprises, the ILO Conventions, and more. In addition, the ten UN Principles of the Global Compact can be specifically mapped to RepRisk’s 28 Issues. Currently RepRisk is in the process of integrating and applying both the SASB Materiality Map® and the UN SDG framework in its risk management and compliance solutions.

Furthermore, RepRisk tags 57 ESG topics and themes (see Figure 10) that are an extension to their core research scope of 28 ESG Issues. ESG Topic Tags are specific and thematic in nature, and one Topic Tag can be linked to multiple ESG Issues. These tags are dynamic, expanding over time in response to client feedback and emerging trends.
In addition to the aforementioned 28 ESG issues and 57 topics tagging, RepRisk also provides a UN Global Compact Violator Flag. The UN Global Compact (UNGC) Violator Flag allows users to identify companies that have a high risk or potential risk of violating one or more of the ten UNGC Principles. With the Flag, it is also possible to see if the UNGC violations are primarily linked to the operations (O) or to the supply chain (S) of a company.

For each company and each Principle, the UNGC Violator Flag has one of three values:

- "Violator": high risk of violating the UNGC Principles
- "Potential": potential risk of violating the UNGC Principles
- "Blank": no strong evidence of violating the UNGC Principles
Companies classified as “UNGC Violators” are those that have had a significant and credible exposure to ESG risk incidents associated with one or more of the ten UNGC Principles.

**Scores**

Each day more than 500,000 risk incidents are pre-selected through advanced text and metadata extraction from unstructured content and undergo multilingual de-duplication and clustering processes. Sentiment and relevancy scoring based on entity identification (e.g. companies, issues or countries) further support the automatic identification of relevant risk incidents.

Each risk incident is automatically linked to all the entities identified in the original source, including the related companies, projects, sectors, countries, ESG Issues etc. Such linking serves to analyse the relationships between various entities and issues on the RepRisk Platform.

If a particular risk incident appears in multiple sources, the incident is accounted only once from the most influential source which is perceived to reflect the overall reputational risk. When the same story appears in multiple languages including English, the English language news source is selected – again, to reflect the reputational risk. The incident is only added again if the risk profile of the incident changes.

The results of the screening process are then passed onto to the analyst team, responsible for reviewing and approving the source selection, automated linking, relevancy scoring, and news analytics according to RepRisk’s proprietary rule-based system. Their analyst team further curates the identified risk incidents and also produces a risk summary.

Each risk incident is analysed according to the following three parameters:

**Severity (harshness) of the risk incident or criticism.** The severity is determined as a function of three dimensions:

- The consequences of the risk incident (For example, with respect to health and safety: no further consequences, injury, death)
- The extent of the impact (one person, a group of people, a large number of people)
- Whether the risk incident was caused by an accident, by negligence, or intent, or even in a systematic way

There are three levels of severity: low severity, severe, and high severity.

**Reach (influence) of the information source, as rated by RepRisk.**

All sources are pre-classified by reach (influence, based on readership/circulation, also a proxy for credibility) into three categories: low influence, medium influence, and high influence.

- Low influence sources consist of local media, smaller NGOs, local governmental bodies, social media etc.
- Medium influence sources include most national and regional media, international NGOs, and state, national, and international governmental bodies.
- High influence sources are the few truly global media such as the NY Times, BBC, South China Morning Post, and others.

**Novelty (newness) of the issues, i.e. whether this is the first time a company is exposed to this issue.**

For each item, a RepRisk Analyst writes an original title and abstract in English that summarises the risk incident. For simple risk incidents where the title captures the relevant issues, an abstract may not be needed.

A risk incident is only entered once onto the RepRisk Platform, unless the risk profile of the incident changes. This can occur in one of the three following scenarios, which increases the reputational risk of the incident for the company concerned:

- Story development: a new development appears related to the same issue.
- Source escalation: the issue appears again in a more influential source.
Six-week rule: the issue appears again for the same company in the same country after a six-week period, which is a potential signal that the issues remain unresolved.

Before an incident is published on RepRisk’s Platform, it needs to undergo a quality assurance check and approval by a senior RepRisk Analyst. This process ensures that the completed analysis has been conducted in accordance with RepRisk’s rule-based methodology.

Risk incidents that are very severe or come from influential sources are processed on the same day or within 24 hours to appear on their platform. Severe incidents are typically processed in 2 days, followed by low-severity incidents which could take one week.

The final step in the process, the quantification of the risk, is done through data science. The RepRisk Index (RRI) dynamically captures and quantifies reputational risk exposure related to ESG issues, and the RepRisk Rating (RRR) utilises a letter-rating system (AAA to D) which facilitates benchmarking and integration of ESG and business conduct risks, as well as other metrics such as the UN Global Compact Violator Flag.

There are proprietary standard and customized risk metrics offered by RepRisk.

**The RepRisk Index (RRI)**

RRI measures reputational risk exposure. It is a proprietary algorithm that dynamically captures and quantifies a company’s or project’s reputational exposure to ESG and business conduct risks. It allows the comparison of a company’s exposure with that of its peers’ and helps track the risk trend over time.

In essence, RRI facilitates an initial assessment of the ESG and business conduct risks associated with financing, investing, or doing business with a particular company.

RRI ranges from 0 (lowest) to 100 (highest). The higher the value, the higher the risk exposure:

- **0-24 low risk exposure**
- **25-49 medium risk exposure**
  - Most large multinationals have RRIs in this range due to their global footprint and salience.
- **50-59 high risk exposure**
- **60-74 very high risk exposure**
- **75-100 extremely high risk exposure**
  - RRI is calibrated in such a way that only a handful of companies that are extremely exposed to serious ESG issues would ever reach this threshold

Within RRI, there are three metrics provided as below:

- **Current RRI** – denotes the current level of media and stakeholder attention of a company related to ESG issues.
- **Peak RRI** – represents the highest level of the RRI over the last two years
- **RRI Change or Trend** – calculated as the increase or decrease of RRI within the past 30 days.

The calculation of the RRI is based on two factors:

- **News Value**: each news story impacts RRI to a different extent depending on the reach, severity and novelty of the news.
- **News Intensity**: frequency and timing of the information.

RRI emphasises companies that are newly exposed or have had less exposure to the ESG issues in the past. It does not rely on the sequence of news nor does it change depending on whether an issue is an E/S/G issue.
There is no specific weighting of the ESG Issues by sector or country. RRI does not provide individual E/S/G components. Instead, they provide a breakdown of each RRI score by the number of associations a company has with the aggregated E/S/G issues, given as percentages. However, this breakdown should not be used for company-to-company comparisons, as it considers only the number of risk incidents per pillar, without any considerations in terms of severity, reach and novelty. Consequently, it should be used for monitoring how a particular company’s E/S/G exposure has developed over time.

For any given day, there are two scenarios that can happen:

- A new risk incident for a company or project has occurred in which case RRI would be recalculated. The magnitude of the change depends on the severity, reach, and novelty of the incident.
- No new risk exposure, in which case RRI would decay.

The decay schedule of RRI over time is as follows:

- For the first 14 days after a significant risk incident, the Current RRI carries the same value.
- If no new exposure is captured thereafter, the Current RRI decays to zero over a maximum period of two years.
  - If the Current RRI is in the range of 25-100 and no significant exposure is captured, it decays at a rate of 25 every two months until it reaches 25.
  - If the Current RRI is at or below 25 and no significant exposure is captured, it decays at a rate of 25 every 18 months until it reaches zero.
Company Examples

Below are two examples provided by RepRisk that demonstrate how their RRI offers insights into companies’ ESG behaviour.

The first example is Pacific Gas & Electric Company (PG&E) which was sued by victims of the historic Camp Fire in California in November 2018. The lawsuits accused the company of failing to maintain proper infrastructure and safety equipment which was suspected to have caused the initial fire. RRI categorised the company as "very high risk exposure" two months before the disaster, with the breakdown highlighting significant environmental and social concerns.

![Figure 11. RepRisk Index Example – PG&E](image)

Source: RepRisk

The second example is Vale SA which was plunged into crisis when the tailings dam of its Corrego do Feijao B1 iron ore mine collapsed near Brumadinho in the Brazilian state of Minas Gerais in January 2019. According to RepRisk’s RRI, the company had a very high ESG exposure level with Peak RRI 62 nearly two years before the incident happened. The high score had been triggered by a series of allegations and risk incidents which raised serious concerns about the business conduct of the company. In the wake of the dam collapse incident, the share price of the company dropped by 20%.

Before the collapse of the dam, Vale had put numerous ESG policies in place, e.g.

- Conducted two ESG seminars in 2018
- Instigated their digital transformation programme for sustainability
- Complied with global ESG reporting initiatives
- Committed to Sustainable Development Goals

Vale was also selected to be a member of Corporate Sustainability Index in Sao Paulo. However, despite all these policies, data collected by RepRisk flagged 590 times that Vale was linked to RepRisk’s ESG issues and 99 times to ESG Topic Tags including impact on communities, landscape, ecosystem, local pollution and waste issues. We can see from the evolution of Vale’s RRI, environmental and social issues have been the main concerns that drive up the ESG risk profile of the company.
As mentioned earlier, RepRisk also provides a UNGC Violater Flag. The underlying risk metric of the Violator Flag is the RepRisk UNGC Violator Index (UNGC VI), which is based on the ESG risk incidents related to a company over the previous two years. Very severe risk incidents are given a higher importance than severe and less severe risk incidents. Furthermore, the UNGC VI underweights risk incidents reported in low reach (influence) sources.

The threshold for being classified as a “UNGC Violator” is higher for highly scrutinized companies, particularly multinationals and conglomerates that are more exposed due to their size, global footprint, and salience in media and stakeholders. This approach helps to balance the information available on smaller companies which may inherently be more vulnerable to risks.

The RepRisk Rating (RRR)

RRR is a letter-rating system (AAA to D) that facilitates corporate benchmarking against its peer group and the sector, as well as integration of ESG and business conduct risks into business processes. It provides decision support in risk management, compliance, investment management, and supplier risk assessment.

In terms of calculations, RRR combines two factors below:

- The company-specific ESG risk exposure, provided by the Peak RRI
- The country-sector ESG risk exposure, provided by the country-sector average of the company, calculated by:
  - The headquarter's ESG Risk Exposure value (weighting 50%) which is the country-sector value of the company’s country of headquarters and primary sector
  - The international ESG Risk Exposure value (weighting 50%) which is the average of all country-sector values of the country-sector combinations where the company has been linked to ESG risk incidents

RepRisk then transforms the two factors collected and arrives at the overall RRR which ranges from AAA to D:

- AAA, AA, A denotes low ESG risk exposure
- BBB, BB, B denotes moderate ESG risk exposure
- CCC, CC, C denotes high ESG risk exposure
- D denotes very high ESG risk exposure
EMPIRICAL ANALYSIS

Coverage

RepRisk states that there are 120,000 companies under coverage on their platform, including both public and private firms. They have created their own sector classification based on the ICB Classification which consists of 34 proprietary sectors. These sectors are then mapped to the 88 European Classification of Economic Activities (NACE)5 divisions for a more granular classification. For our assessment, we examine their stock coverage using MSCI World and Emerging Market index universes and GICS for sector analysis.

As mentioned previously, RepRisk is designed to flag risk incidents and violations in relation to the ESG issues/topics and hence the focus is on reputational risk. Based on the decay schedule described in the earlier section, if a company had a risk incident before but did not deteriorate further or have any new incidents, its RRI would decay to zero within a maximum timeframe of two years, and then become NA if no more risk incidents are flagged. However, it does not mean the company is no longer 'covered' as RepRisk continues to monitor risk incidents through the data sources they employ.

With standard coverage checks, i.e. how many stocks carry scores at time t compared to our universes (World and EM), such assessments could be misleading in RepRisk's case as the fact that companies' current RRIs move from zeros to NAs does not mean they are dropped out of RepRisk's coverage. Consequently, the two charts below show the companies RepRisk monitor rather than those that have non-NA scores.

It can be seen from below that the coverage of MSCI World has been consistently good, especially over the last 5 years, achieving over 90% of the universe. Emerging markets are more challenging given the language barriers and also perhaps lower frequency of ESG-related news in the media and datafeeds in these markets. Despite these issues, RepRisk has been covering over 70% of the MSCI EM stock universe since 2010.

To check if the data has pronounced size biases, in the above charts we analyse the coverage by market cap buckets for both MSCI World and EM universes. As noted earlier, with RepRisk's construction of their Current RRI, where only companies that have had risk incidents over the last two years would carry non-NA scores, the coverage shown here is for companies with flagged risk incidents rather than general data coverage, as these data points form the basis of our tilt and performance assessments later on. With that caveat in mind, it is evident from the charts that RepRisk is able to capture risk incidents across market cap buckets in both World and EM. One could perhaps argue that the coverage of the smallest market cap bucket in EM is low and this could be down to genuine lack of risk incidents or lack of data sources for the smallest companies in EM or perhaps both.

---

5 NACE is the Industry Standard Classification System used within EU.
Slicing the data another way, we examine the coverage through time both by market-cap weighting and by number of companies (effectively equal-weighting applied) by sector. By comparing the percentages of market-cap weighted and equally weighted sector coverage, we can see that larger companies do have higher percentages of risk incidents across many sectors, e.g. 9 out of 11 sectors have higher than 85% of companies (market-cap weighted) carry non-NA scores in 2019 while the equally weighted percentages are markedly lower for most of sectors.

For emerging markets, as highlighted earlier, the coverage is not as high as that in developed world, which can be due to genuine lack of incidents or lack of data sources on ESG issues, or perhaps the ESG issues are not high up on the agenda in these markets and thus infrequently reported. On a market-cap weighted basis, the latest percentages of companies with flagged risk incidents (and thus carry non-NA scores) across most sectors are above 50%, while on an equal-weighted basis the percentages are notably lower.
Figure 18. RepRisk Coverage through Time by Sector - EM

<table>
<thead>
<tr>
<th>% of Companies Covered Market-cap</th>
<th>Industrials</th>
<th>Utilities</th>
<th>Financials</th>
<th>Health Care</th>
<th>Materials</th>
<th>Consumer Staples</th>
<th>Consumer Discretionary</th>
<th>Information Technology</th>
<th>Communication Services</th>
<th>Energy</th>
<th>Real Estate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>48%</td>
<td>59%</td>
<td>35%</td>
<td>13%</td>
<td>51%</td>
<td>39%</td>
<td>47%</td>
<td>61%</td>
<td>44%</td>
<td>61%</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>56%</td>
<td>66%</td>
<td>38%</td>
<td>14%</td>
<td>55%</td>
<td>47%</td>
<td>49%</td>
<td>66%</td>
<td>51%</td>
<td>65%</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>53%</td>
<td>70%</td>
<td>42%</td>
<td>13%</td>
<td>55%</td>
<td>44%</td>
<td>58%</td>
<td>71%</td>
<td>61%</td>
<td>66%</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>58%</td>
<td>70%</td>
<td>49%</td>
<td>25%</td>
<td>60%</td>
<td>56%</td>
<td>61%</td>
<td>76%</td>
<td>62%</td>
<td>65%</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>57%</td>
<td>65%</td>
<td>48%</td>
<td>33%</td>
<td>63%</td>
<td>59%</td>
<td>59%</td>
<td>76%</td>
<td>65%</td>
<td>64%</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>60%</td>
<td>68%</td>
<td>47%</td>
<td>33%</td>
<td>68%</td>
<td>59%</td>
<td>60%</td>
<td>70%</td>
<td>67%</td>
<td>63%</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>64%</td>
<td>67%</td>
<td>51%</td>
<td>47%</td>
<td>72%</td>
<td>62%</td>
<td>59%</td>
<td>68%</td>
<td>67%</td>
<td>59%</td>
<td>50%</td>
</tr>
<tr>
<td>2018</td>
<td>65%</td>
<td>69%</td>
<td>52%</td>
<td>59%</td>
<td>68%</td>
<td>62%</td>
<td>69%</td>
<td>68%</td>
<td>63%</td>
<td>56%</td>
<td>63%</td>
</tr>
<tr>
<td>2019</td>
<td>63%</td>
<td>63%</td>
<td>50%</td>
<td>49%</td>
<td>66%</td>
<td>56%</td>
<td>37%</td>
<td>77%</td>
<td>76%</td>
<td>52%</td>
<td>63%</td>
</tr>
</tbody>
</table>

Source: RepRisk

Tilts

RepRisk provides a breakdown in terms of the percentage of the number of associations a company has within the aggregated E/S/G issues. While not suitable for company-to-company comparisons, the average E/S/G risk incident percentage breakdowns over time (see Figure 19 and Figure 20) paint an interesting picture in that governance issues have become much more dominant over the past five years or so, accounting for more than 50% of the risk incidents regardless of being in the developed world or emerging markets.

From a sector perspective, Figure 21 and Figure 22 reveal the varying importance of E/S/G components per sector as shown overleaf. Consistently across markets, Governance issues are particularly important in Communication Services, Financials and Healthcare sectors, while Environmental issues are the key focus in Energy, Materials and Utility sectors. One sector that appears to have different E/S/G emphasis between World and EM is Information Technology, where developed world Governance is the most important issue while in EM the Social aspect appears to matter more. This might be down to the degree of maturity of the sector in different markets for example, in EM labour rights are less protected than in the developed world and hence still dominate the Social agenda.

Figure 19. RepRisk E/S/G Component Breakdown through Time: World

Figure 20. RepRisk E/S/G Component Breakdown through Time: EM

Source: RepRisk, CGDI, MSCI

Source: RepRisk, CGDI, MSCI
Next, we turn our attention to the RRI distribution by sector for both the developed world and emerging markets. This analysis can tell us whether or not there are differences between sectors in terms of the mean values and also the one standard deviation range.

The charts below indicates that the Energy sector tends to have higher RRIs in both universes, while the Real Estate sector appears to have the lowest RRIs, although the +/- one standard deviation ranges are similar across sectors. This suggests the Energy sector carries higher risks, relative to Real Estate sector.

**Figure 23. RepRisk Sector Mean and +/- Standard Deviation: World**

**Figure 24. RepRisk Sector Mean and +/- Standard Deviation: EM**

**Performance**

In terms of performance where we link ESG performance with stock returns, we use a quintile approach to examine the basket returns (equally weighted) based on the current RRI values. Note that higher RRIs indicate higher risk and thus our quintiles are formed in a reversed order where the best quintile, Quintile 5, captures stocks with the lowest current RRIs (including zeros but excluding NAs) and vice versa.

The charts in Figure 25 show that Quintile 5 has outperformed Quintile 1 since January 2010 but one can clearly see that the outperformance is mainly driven by Quintile 1’s underperformance, rather than the top quintile beating other quintiles or our benchmarks (Figure 26) of MSCI World regardless of being market-cap weighted or equal-weighted and ESG Leaders indices.
This demonstrates RepRisk Current RRIs are a powerful tool to identify underperformers, not only based on the ESG issues highlighted in the Methodology section but also in their financial performance as those stocks attract lower returns.

**Figure 25. RepRisk Quintile Performance based on Current RRI: World**

**Figure 26. RepRisk Top/Bottom Quintile Performance vs Benchmarks**

**Figure 27. RepRisk Quintile Performance Statistics vs MSCI World Indices**

<table>
<thead>
<tr>
<th></th>
<th>Quintile 1</th>
<th>Quintile 5</th>
<th>World ESG Leaders</th>
<th>World (Cap-weighted)</th>
<th>World (Equal-weighted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annualised Returns</td>
<td>7%</td>
<td>10%</td>
<td>10%</td>
<td>10%</td>
<td>9%</td>
</tr>
<tr>
<td>Annualised Volatility</td>
<td>15%</td>
<td>14%</td>
<td>13%</td>
<td>13%</td>
<td>14%</td>
</tr>
<tr>
<td>Sharpe Ratio</td>
<td>0.48</td>
<td>0.72</td>
<td>0.75</td>
<td>0.76</td>
<td>0.68</td>
</tr>
</tbody>
</table>

**Source:** RepRisk, CGDI, MSCI

On a long-short basis, the performance based on (Quintile 5 – Quintile 1) yields an annualised return of 2.4% with volatility of 4.3% giving a moderate Sharpe Ratio of 0.55.

**Figure 28. RepRisk Current RRI Long/Short Performance: World**

**Source:** RepRisk, CGDI
Utilising the Fama-French-Carhart regression⁶, it shows that the long-short return has significant growth, small caps and price momentum tilts and the residual return after accounting for these exposures is statistically significant at the 90% level. We also check the return correlations with the well-known style factors / risk premia. The table above suggests that the long-short returns have growth, small caps, high price momentum and lower risk tilts which are in stark contrast to the tilts in analyst-driven ESG research where typically carry positive value with little momentum exposures.

When applying the same methodology to assess the performance based on RepRisk’s Current RRI in emerging markets, the efficacy of the strategy is less than satisfactory where there is little difference between Quintile 5 and Quintile 1 and both track the MSCI EM benchmarks but substantially underperform MSCI EM ESG leaders (see Figure 30 below). This could be seen as one of the main shortcomings where the risk incidents are not frequently reported and thus not enough creditable sources could be used for ESG assessments on companies using AI / machine learning techniques. In this case, an analyst-driven ESG valuation could be more fruitful even though the update frequency is typically annual.

---

⁶ We use pure style indices constructed by Citi Quant Research team to represent the value, size, and momentum premium.
SUMMARY

In this section we discuss the philosophy, construction and methodology of RepRisk's core offerings. Specifically, we focused on how their ESG criteria are selected and applied in their products with a view to unveiling the AI and machine learning / NLP capabilities they employ to deliver an outside-in perspective on companies. In addition to the detailed description of product construction, we also examine their offerings through a practical lens in terms of how investor clients can potentially use their data to guide their ESG investing processes.

In order to do so, we have used MSCI World and Emerging Markets indices as the universes and discuss data coverage, tilts and performances using RepRisk's data. An interesting finding through our tilt analysis is that Governance has been growing in terms of importance within E vs S vs G. Also our sector breakdown analysis demonstrates the differences in terms of E/S/G focus across sectors. Lastly, we look at how effective their Current RRI could be used to identify high ESG risk firms and link them with stock returns.

Our results presented in earlier sections show that the firms categorised as higher risk by RepRisk indeed are associated with lower returns in the developed world, using a simple quintile construction where we compare the returns from the top quintile (lowest risk) with the bottom quintile (highest risk). The positive return spread between the top quintile and the bottom quintile was predominantly driven by the underperformance of the bottom quintile, which is what RepRisk set out to do. Using the Fama-French-Carhart framework, we are able to show that after accounting for the market, size, value and momentum factors, the residual return is still statistically significant at the 90% level which shows the value-add RepRisk potentially has to investment returns.

While the results from developed markets are encouraging, the same can't be said about emerging markets and that can be attributed to a number of reasons such as lack of credible data sources to conduct proper outside-in assessments. This could be down to lack of reported issues by the media which could indicate lack of attention on ESG issues in the local market or genuinely no risk incidents worth flagging. However, the strong outperformance of the MSCI EM ESG Leader index would suggest the former is more likely to be true.

Based on our analysis, we have established that RepRisk is an effective ESG tool that alerts and flags ESG risk incidents of companies which have evidently attracted lower returns. However, the applications of ESG in the investment management industry have expanded beyond negative screening / exclusions to include positive / best-in-class and full ESG integration into portfolios. With the current focus being on the negative side, RepRisk might not be as helpful in assessing the positive ESG aspects of companies apart from implicitly inferred by the lack of risk incidents.
Truvalue Labs was founded in 2013 and strives to be the market leader in applying AI and big data analytics to score companies' ESG behaviours as they happen, rather than how firms report them to be. Truvalue Labs' solution delivers objective algorithmic scoring on ESG factors as identified by the Sustainability Accounting Standards Board™ (SASB™) as having a material impact on company value by industry and sector. SASB aims to meet the need for industry-specific reporting standards, for the ease of comparisons and benchmarking.

Truvalue Labs uses artificial intelligence to mine massive amounts of unstructured external data to uncover not just risks, but also opportunities created by positive ESG behaviour of companies. They cover 15,000+ publicly listed companies globally and capture over 100,000 external sources such as local and international news, industry publications, and NGOs to report evidence of ESG performance across a range of financially material factors, delivering insight into behaviour companies may choose not to report.

In order to provide such insight, Truvalue Labs has developed their proprietary AI to unlock ESG insights hidden in massive amounts of unstructured data, rather than sourcing from other existing ESG providers. Consequently, they are able to offer users the capability to drill down and examine every piece of information for any events that move overall scores within particular ESG categories.

**Figure 31. Truvalue Lab’s Score Generation Flow**

**TVL algorithms identify and categorize material issues, quantify them, then produce a suite of data and analytics**
METHODOLOGY

Philosophy

Truvalue Labs believe that technology can uncover new insights at a speed and scale which is never before possible. Their mission is to deliver scalable, consistent, and transparent ESG solutions that can disrupt and transform an industry and challenge the status quo. The company believes their methodology achieves scalability by building systems that keep up with the rapid expansion of the world’s data. In order to ensure consistency, Truvalue leverages technology to score companies instead of relying on individual analysts to provide impartial assessments. Furthermore, the company emphasizes transparency in that the data used to derive the scores is always visible to users for validation or further research purposes.

Truvalue Labs believe that although the data era necessitates the application of artificial intelligence, human intelligence also plays an essential role in defining the analytical framework in which machine-driven analysis is guided and calibrated. To harness the power of technology for discerning ESG insights from unstructured data, human intelligence is applied at the beginning of the process rather than at the end, as in the traditional ESG research model.

With their AI framework, human ESG domain expertise is embedded in the model from the outset which helps generate consistent results at the end of process. This enables investment professionals to spend more time on the analysis that matters the most in achieving their investment outcomes.

As Truvalue Labs believe that ESG factors can have a material impact on financial performance, their ESG dataset therefore consists of a set of characteristics that describe intangible risks and opportunities upon which investors can make better decisions to achieve greater risk-adjusted returns over the short, medium and long-term.

Sources of Data and Update Frequency

Everyday Truvalue Labs ingest data drawn from more than 100,000 sources which are mined for information relevant to ESG-related company activities using AI, applying a consistent, systematic framework for identification and analysis at scale. The sources include local, national, and international news, reports by NGOs, watchdog groups, trade blogs, industry publications, and important events shared by industry thought leaders via social media platforms such as Twitter.

All data is analysed with a consistent process by algorithms. Data is updated on a continuous rolling basis throughout each day. At the time of writing, Truvalue Labs score events in English only, but their upcoming July release will introduce coverage in seven languages: English, German, Spanish, French, Japanese, Italian, and Portuguese, which should broaden their coverage and address the languages issues in emerging markets.

How AI / Machine Learning is Applied

The AI technology behind Truvalue Labs’ offerings augments human decision making by extracting meaningful sustainability signals from large volume of unstructured data. Truvalue Labs support on-demand analytics and provide users with instant access to original data sources on their platform. Figure 32 depicts the steps undertaken by Truvalue to derive their scores using AI.
Truvalue Labs’ Application of AI

Step 1. Collects unstructured data from more than 100,000 sources

Truvalue Labs aggregate a variety of data sources into a continuous stream of relevant ESG data for monitored companies and sectors. The scalable nature of the technology potentially allows limitless expansion of sources. Their framework evaluates both semantic and quantitative content, and the flexible architecture allows users to incorporate their own proprietary data sources.

Step 2. Extracts relevant metrics

Truvalue Labs sort content flow by data type, then establish the context around each data point to extract and categorise content based on sustainability. Items are then tagged to multiple categories if a particular data point has relevance to each. In the SASB materiality view, the company’s score charts are made up of data in a subset of the categories that SASB’s research shows to be financially material for that company, based on its industry. The flexibility of the platform enables the application of multiple “lenses”.

Step 3. Normalizes data and generates sustainability performance analytics

All monitored companies have dynamic scorecards that display their ESG trends. To arrive at the company’s score, each data point is weighted according to the timeliness and intensity in scoring formulas that reveal short-term and long-term on the ground performance.
ESG Criteria Used

According to Truvalue, they were the first to the market with a product that employed the SASB’s framework, which allows investment decisions to incorporate performance in line with generally accepted, financially material ESG factors. Since then, its team has mapped SASB’s categories to the United Nations Sustainable Development Goals (SDGs) in a collaborative work between Truvalue Labs and academics from the University of Oxford and the University of Siena. The flexibility of their platform enables the application of multiple frameworks or “lenses” e.g. the UN Global Compact, or a custom framework.

Companies are scored across all SASB categories on a daily basis, as well as on the overall company level. Users can combine any of the material categories to create their own specific scores. This enables tracking of company performance on SDG-related material and non-material ESG issues in a way that enables stock selection and engagement and screening opportunities. The system also allows for identification of public policy gaps and impact investing opportunities by combining a measure of the “scope” of the SDG impact of an industry with the SDG-linked ESG performance scores.

Source: SASB
Scores

Truvalue Labs produce multiple ESG scores for all 15,000+ companies tracked within its universe, to provide users with a view of companies’ short, medium and long-term ESG performance in order to meet the various investment objectives associated with different strategies and asset classes.

- **Pulse**

  The Pulse Score reveals near-term performance changes on a real-time basis, which can be both positive and negative changes in behaviour associated with day-to-day events and activities. It is the most real-time view of ESG performance available from Truvalue and enables users to set up monitoring and alerts in order to be notified of events which could impact the performance of a company or portfolio. While a single pulse change may not translate into an enduring effect on longer-term financial performance, frequent spikes on a given topic can signal an emerging risk (or improvement) which may not have been widely identified but may warrant deeper investigations.

- **Insight**

  The Insight Score is a longer-term measure, akin to an ESG rating. It is a more moderate measure of performance, providing an overall view of how a company has been performing on a rolling 6-month basis. It facilitates meaningful comparisons between different companies, portfolios and sectors, and can therefore be deployed in screening strategies by identifying the best or worst performers, or on a single entity basis by comparison to an industry, benchmark or peers. Insight scores can be applied on a portfolio, company and individual category levels, allowing for deeper macro and fundamental analysis.

- **Momentum**

  The ESG Momentum Score measures the trend of a company's Insight score. ESG Momentum scores are used to indicate whether performance is generally improving, or deteriorating. Based on a rolling twelve-month period, it is an ESG metric that captures the long-term pattern of corporate behaviour, allowing a meaningful comparison of performance versus peers.

- **Volume**

  The Volume Score tracks the number of articles related to each company over a trailing twelve-month period. This measure provides users with transparency into the amount of underlying content that is driving each company’s ESG scores. Within the Truvalue Platform, the volume of company events is presented as one of the following: High, Medium, Low, or No Data. This score is displayed using a three-bar scale and can be found on the Company, Portfolio, Industry and Sector views of Truvalue Lab’s Platform. Their Data Services clients can have direct access to volume counts daily, as well as on a trailing twelve-month basis.
An additional offering by Truvalue Labs is an enhanced view of the material ESG factors impacting a particular company, portfolio, industry, benchmark, or sector.

As they fully integrate SASB’s materiality mapping framework within their dataset, users can choose to filter by SASB materiality, or evaluate materiality in other ways. One of the important characteristics of materiality frameworks is that they can be applied on an industry and also on an entity level, through a 5th score type which is the Category Impact %.

The Category Impact % reveals the specific categories that drive a company’s overall score over a trailing twelve-month period. Higher Category Impact % values indicate higher volumes of information flow in the market, related to specific categories and reflect the ESG categories deemed important to the outside world.

**Company Examples**

In the following page, we show two stock examples that are provided by Truvalue which demonstrate how their scores are able to pick up deterioration of ESG performance and how it has correlated with share price underperformance.

Throughout 2018, Facebook was criticised for the company’s handling of user data, fake news and GDPR issues, and there had been calls for regulations on social media providers. These negative headlines not only dented the earnings outlook of Facebook but ESG concerns, particularly on governance were also reflected in Truvalue’s Pulse scores, decreasing by more than 21% in 2018. The drivers of the deteriorating scores also included a security breach where nearly 50 million users were affected, increasing both European and US scrutiny over the company’s conduct. Facebook’s share price and Pulse score have since co-moved closely together.
The second example is Monsanto which was bought mid-year last year by Bayer. Monsanto suffered considerable litigation risk in 2018 and in particular faced a cancer-related lawsuit brought by a school groundskeeper who sued the company for causing him to develop terminal cancer due to exposure to their Roundup weed killer product in the course of his job. The Superior Court of California in San Francisco awarded the groundskeeper $289 million in damages. Truvalue's Pulse scores started to fall in spring and accelerated in summer as there were over 515 cancer lawsuits pending, together with accusations of collusion with the EPA. These litigation risks weighed on Bayer's share price and continued to intensify, causing the Pulse score for Bayer to drop dramatically in early October 2018, two weeks before the share price slumped on 22 October as a judge reaffirmed the verdict with regard to the Roundup lawsuit, although slashing the payout to $78 million.
EMPIRICAL ANALYSIS

Coverage

The systematic, technology-driven way in which Truvalue Labs produces their scores means they are able to deliver consistent historical data for backtesting performance purposes with the history going back to 2007.

Another data detail worth mentioning is regarding their scalability; Truvalue Labs employ multi-pipeline architecture so that new products do not disrupt existing data sets users rely on. The flexibility of the platform enables the applications of multiple frameworks or “lenses”. This enables them to offer a framework agnostic product suite where it is possible to map the data to SASB, UN SDGs, or any other proprietary ESG framework and provide consistent history to any framework chosen.

In terms of overall coverage, the current version of Truvalue Labs’ data we have tested cover over 9,000 companies. The new version due out in July 2019 will expand the universe to over 15,000 companies. With the existing universe, we compare the stocks covered by Truvalue’s data for both World and EM universes. As can be seen in Figure 10, the data coverage has improved steadily since 2010 before reaching over 90% from 2015 in developed world universe, while for EM the coverage was lower but still relatively substantive at around 80% over the last five years, considering the language they cover with our testing dataset in EM is only English-based. The English-only drawback became apparent when MSCI started to include China A-shares in May 2018 contributing to the drop in coverage. With the new version to be released in July 2019, the languages covered will expand to seven including many that are essential for assessing emerging markets.

From a market cap coverage perspective, we compare the stocks Truvalue’s data covers for both World and EM universes. The two charts below show the coverage by market cap bucket as of end of December 2018 and the World chart reaffirms the comprehensive coverage Truvalue has in this region, close to 100% in every market cap bucket. For emerging markets, it is clear that the coverage for large caps is substantially better than the lowest market cap bucket.
Another check we perform is on sector coverage as there might be biases introduced by the data sources, e.g. the fact that media covers certain sectors more than the others. In the tables below, we look at the sector coverage both by market cap weighting and also by the number of companies covered for both World and EM universes.

Once again for World universe, it is impressive that the coverage in terms of market cap covered reaches well above 90% for most of the sectors as of 2019 and hardly any sector dropped below 90% prior to that. Analysing the number of companies covered shows that the coverage prior to 2015 was not as good but has drastically improved since.

Turning our attention to EM, as highlighted earlier, the current version we have tested contains data from only English-based sources in these markets. This puts a limitation on their coverage by sector as well, as depicted in the tables overleaf. Regardless of market-cap weighted or by count, the coverage is lower than that of World but most of them are still above 75%, apart from 2019 due to the inclusion of China A Shares.
Figure 42. Truvalue Coverage through Time by Sector – EM (lower than 80% is highlighted in orange)

<table>
<thead>
<tr>
<th>% of Companies Covered</th>
<th>Industrials</th>
<th>Utilities</th>
<th>Financials</th>
<th>Health Care</th>
<th>Materials</th>
<th>Consumer Staples</th>
<th>Consumer Discretionary</th>
<th>Information Technology</th>
<th>Communication Services</th>
<th>Energy</th>
<th>Real Estate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>62%</td>
<td>71%</td>
<td>61%</td>
<td>70%</td>
<td>74%</td>
<td>55%</td>
<td>57%</td>
<td>89%</td>
<td>86%</td>
<td>90%</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>60%</td>
<td>71%</td>
<td>67%</td>
<td>83%</td>
<td>74%</td>
<td>59%</td>
<td>61%</td>
<td>91%</td>
<td>91%</td>
<td>91%</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>66%</td>
<td>76%</td>
<td>72%</td>
<td>79%</td>
<td>82%</td>
<td>64%</td>
<td>71%</td>
<td>93%</td>
<td>91%</td>
<td>91%</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>68%</td>
<td>86%</td>
<td>74%</td>
<td>86%</td>
<td>85%</td>
<td>78%</td>
<td>74%</td>
<td>94%</td>
<td>93%</td>
<td>95%</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>81%</td>
<td>90%</td>
<td>81%</td>
<td>93%</td>
<td>91%</td>
<td>88%</td>
<td>77%</td>
<td>95%</td>
<td>93%</td>
<td>95%</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>86%</td>
<td>91%</td>
<td>90%</td>
<td>96%</td>
<td>97%</td>
<td>94%</td>
<td>87%</td>
<td>99%</td>
<td>96%</td>
<td>99%</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>92%</td>
<td>94%</td>
<td>96%</td>
<td>98%</td>
<td>98%</td>
<td>95%</td>
<td>90%</td>
<td>100%</td>
<td>96%</td>
<td>100%</td>
<td>79%</td>
</tr>
<tr>
<td>2018</td>
<td>91%</td>
<td>93%</td>
<td>98%</td>
<td>90%</td>
<td>96%</td>
<td>92%</td>
<td>96%</td>
<td>98%</td>
<td>94%</td>
<td>99%</td>
<td>82%</td>
</tr>
<tr>
<td>2019</td>
<td>84%</td>
<td>90%</td>
<td>98%</td>
<td>79%</td>
<td>92%</td>
<td>88%</td>
<td>92%</td>
<td>97%</td>
<td>96%</td>
<td>99%</td>
<td>78%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>% of Companies Covered</th>
<th>Industrials</th>
<th>Utilities</th>
<th>Financials</th>
<th>Health Care</th>
<th>Materials</th>
<th>Consumer Staples</th>
<th>Consumer Discretionary</th>
<th>Information Technology</th>
<th>Communication Services</th>
<th>Energy</th>
<th>Real Estate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covered by Count</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>55%</td>
<td>65%</td>
<td>44%</td>
<td>61%</td>
<td>52%</td>
<td>55%</td>
<td>36%</td>
<td>66%</td>
<td>62%</td>
<td>73%</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>57%</td>
<td>66%</td>
<td>49%</td>
<td>75%</td>
<td>55%</td>
<td>57%</td>
<td>42%</td>
<td>68%</td>
<td>76%</td>
<td>76%</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>64%</td>
<td>69%</td>
<td>54%</td>
<td>71%</td>
<td>66%</td>
<td>62%</td>
<td>50%</td>
<td>79%</td>
<td>81%</td>
<td>81%</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>66%</td>
<td>81%</td>
<td>61%</td>
<td>78%</td>
<td>73%</td>
<td>69%</td>
<td>48%</td>
<td>72%</td>
<td>81%</td>
<td>81%</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>79%</td>
<td>86%</td>
<td>71%</td>
<td>84%</td>
<td>83%</td>
<td>84%</td>
<td>58%</td>
<td>81%</td>
<td>87%</td>
<td>91%</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>81%</td>
<td>88%</td>
<td>80%</td>
<td>91%</td>
<td>95%</td>
<td>86%</td>
<td>78%</td>
<td>92%</td>
<td>91%</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>90%</td>
<td>94%</td>
<td>91%</td>
<td>95%</td>
<td>95%</td>
<td>91%</td>
<td>85%</td>
<td>96%</td>
<td>91%</td>
<td>98%</td>
<td>67%</td>
</tr>
<tr>
<td>2018</td>
<td>88%</td>
<td>93%</td>
<td>90%</td>
<td>88%</td>
<td>92%</td>
<td>85%</td>
<td>90%</td>
<td>87%</td>
<td>88%</td>
<td>96%</td>
<td>68%</td>
</tr>
<tr>
<td>2019</td>
<td>57%</td>
<td>72%</td>
<td>70%</td>
<td>54%</td>
<td>67%</td>
<td>68%</td>
<td>63%</td>
<td>60%</td>
<td>74%</td>
<td>85%</td>
<td>47%</td>
</tr>
</tbody>
</table>

Source: Truvalue Labs, CGDI, MSCI
Tilts

As discussed earlier, Truvalue offers four related ESG scores: Pulse, Insight, Momentum and TTM Volume. Pulse is the fastest signal as it measures the day-to-day ESG performance of the companies and Insight is the rolling EWMA 6-month Pulse akin to ESG ratings. In our tilts analysis, we focus on these two measures as they are suited for asset managers who have shorter-term and medium-term investment horizons respectively.

In Figure 43 and Figure 44, we show the mean and +/- one standard deviation by sector for both Pulse and Insight scores in MSCI World and EM universes. Generally speaking, IT sector has the highest Pulse and Insight scores in developed markets while Health Care, Industrials and Real Estate claim the top mean scores in emerging markets.

In the developed world, the means are consistent between Pulse and Insight scores from a sector perspective but the standard deviations are evidently larger in Pulse which should not be surprising due to the short-term nature of the score. Interestingly the opposite is true for EM which could be seen as an indication of lack of newsflows / datafeeds in these markets that are in English on a day-to-day basis. This demonstrates the need for broader language coverage to address the shortcoming which is what Truvalue looks to improve in their new release next month.

![Figure 43. Truvalue Pulse Score Distribution by Sector: World](image)

![Figure 44. Truvalue Insight Score Distribution: World](image)

![Figure 45. Truvalue Pulse Score Distribution by Sector: EM](image)

![Figure 46. Truvalue Insight Score Distribution: EM](image)
Performance

Focusing on quintile performance based on their Insight scores generated with Truvalue’s current version of data, Figure 47 shows that there is a sizeable return spread between Quintile 5 and Quintile 1 although the return profile across all quintiles is not monotonic. When compared to MSCI World benchmarks, Quintile 5 does outperform all of them although the magnitude of outperformance is moderate. Nevertheless, the charts below demonstrate that Truvalue’s Insight scores can be used to select outperforming stocks especially since 2016.

As discussed earlier, Truvalue also generates Momentum scores which are the slope of Insight Scores over the last twelve months, capturing the trend of ESG performance of companies. Their volume scores which are based on article volumes could serve as a useful gauge of how much data is used behind the Insight score. To that end, we construct a simple strategy where we condition the Insight score with strong momentum and good article volumes.
Specifically, we explore two strategies – Long 1 and Long 2 – where we take the top tertile and top quintile respectively based on Insight scores and conditioned the companies in the bucket on having top momentum and good article volumes. Due to the intersection nature of the construct, we check the number of holdings in these two strategies to ensure these ESG portfolios are not overly concentrated. In fact, the average number of holdings in Long 1 portfolio is 53 while that in Long 2 portfolio is 26. As can be seen on the previous page, both Long 1 and 2 substantially have outperformed the benchmarks, delivering over 2% and 3% more returns per annum respectively, relative to the benchmarks. Even though these strategies do carry higher volatility, on a risk-adjusted basis, Long 2 portfolio still outperforms all three benchmarks while Long 1 portfolio beats MSCI World (equally weighted).

Dissecting the performance using the Fama-French-Carhart regression shows that the return spread between top quintile and bottom quintile based on Insight scores can be largely attributed to higher beta and higher momentum tilts with muted value exposure. In contrast, Long 2 portfolio has statistically significant tilts towards Value and Price Momentum with muted beta exposure and the residual return is not statistically significant at the 90% level. This suggests that the combination of Insight, Momentum and Volume scores has mitigated the market beta tilt and elevated the value exposure which appear to have contributed to the strong performance we have seen here.

Using the same evaluation methodology in EM, the results however are not encouraging. The quintiles in fact show performance that is far from monotonic where quintile 3 outperforms all the other quintiles with quintile 4 & 5 at the bottom. Compared to MSCI benchmarks, the performance of Quintile 5 appears to be on par with the broad market but substantially underperforms MSCI EM ESG Leaders index. This could be down to the lack of data sources as highlighted earlier and the fact that the version of the data Truvalue provided us with only covers English-based feeds.

Despite the shortcomings within the dataset, we examine whether using Insight scores conditioned by momentum and article volumes could also uplift the performance compared to using Insight scores alone. In Figure 53, we show the result of conditioning Insight scores with Momentum and Volumes. The construction of the "Long" portfolio is exactly the same as per the "Long 2" portfolio set-up in the developed world. While it does indeed outperform the generic quintiles based on just Insight scores, the Long portfolio still fails to beat the MSCI EM ESG Leaders index. The return profile also looks rather volatile which is due to the lower number of names covered in the universe, coupled with conditionality imposed, resulting in a smaller number of companies captured.

\*Top momentum is defined as top quartile based on Truvalue Momentum scores and the top half of universe based on article volumes.

---

**Figure 50. Truvalue Strategy Performance Statistics – World**

<table>
<thead>
<tr>
<th></th>
<th>Quintile 1</th>
<th>Long 1</th>
<th>Long 2</th>
<th>World ESG Leaders</th>
<th>World (Cap-weighted)</th>
<th>World (Equal-weighted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annualised Returns</td>
<td>8%</td>
<td>12%</td>
<td>13%</td>
<td>10%</td>
<td>10%</td>
<td>9%</td>
</tr>
<tr>
<td>Annualised Volatility</td>
<td>14%</td>
<td>16%</td>
<td>16%</td>
<td>13%</td>
<td>13%</td>
<td>14%</td>
</tr>
<tr>
<td>Sharpe Ratio</td>
<td>0.59</td>
<td>0.74</td>
<td>0.80</td>
<td>0.75</td>
<td>0.76</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Source: Truvalue Labs, CGDI, MSCI

**Figure 51. Truvalue Insight Quintile Performance: EM**

**Figure 52. Truvalue Insight Top/Bottom Quintile Performance vs EM benchmarks**

Source: Truvalue Labs, CGDI, MSCI

---

38 | The Rise of AI in ESG Evaluation
in the portfolio. We also explore whether the opposite is true if we screen for companies with poor Insight scores, deteriorating momentum but backed by good article volumes. The short portfolio in the chart below demonstrates that the strategy does capture companies that underperform quite well, although the same higher volatility issue and lower counts of companies are present yet again. This indicates such strategies might be better suited to fundamental stock screening.

Figure 53. Truvalue Insight conditioned by Momentum and Volumes - EM

![Chart showing performance of short and long portfolios conditioned by momentum and volumes.](chart.png)

Source: Truvalue Labs, CGDI, MSCI

---

8 Average number of holdings is 21.
SUMMARY

In this section, we presented Truvalue Labs' philosophy, methodology, how through AI / machine learning they incorporate ESG criteria into their scores, followed by empirical analysis on data coverage, tilts and performance. Truvalue Labs' scores are designed to capture both positive and negative aspects of companies' behaviour from an ESG perspective.

Utilising over 100,000 data sources, Truvalue Labs applies AI algorithms to score companies based on SASB Sustainability categories on a daily basis to derive their Pulse, Insight, Momentum and Volume scores where the first three scores are designed to suit different investment horizons. Pulse is their near real-time assessment of companies while Insight is a longer-term measure which is the exponentially weighted moving average of Pulse, comparable to ESG scores provided by other vendors. Momentum is the long-term metric as it is the slope of Insight scores from the trailing twelve months, indicating the trend of corporate behaviour.

In terms of coverage, Truvalue Labs has achieved over 90% of developed markets as defined by MSCI World universe since 2015 but emerging markets are less well-covered. The main contributing factor is language as the dataset we use for the study contains only English-based data sources which limit the coverage to a large extent in EM.

Performance-wise for developed markets, we have demonstrated that using Truvalue Labs' Insight scores in identifying good vs poor ESG performance companies does deliver alpha. The outperformance could be further enhanced when Insight scores are combined with positive momentum and decent article volumes. For emerging markets, the results were not satisfactory due to limitations on coverage and language. However, repeating the same combination of good Insight scores with strong positive Momentum and sizeable article Volume shows performance does improve while the opposite is true – companies with poor Insight scores combined with deteriorating momentum backed by decent article volume yields worse returns in all cases examined.

Overall, the framework Truvalue Labs employs to evaluate companies from an ESG perspective is interesting as the scalability of their infrastructure makes expansion of data sources and version controls straightforward. The limitations of the current dataset in terms of coverage and language especially in emerging markets should be addressed in their upcoming data release in July when Truvalue Labs broaden their coverage to 15,000 companies and covers seven languages.
The Methodology and Construction sections have benefitted from discussions with Arabesque S-Ray, with thanks to Andreas Feiner, CEO, andreas.feiner@arabesque.com, Alexandra Pavlovskis, Alexandra.Pavlovskis@arabesque.com and team.

Founded in 2013, Arabesque takes a big data approach in quantifying ESG topics by aggregating ESG data from a wide-range of third party data sources into a single company-level score. S-Ray is the tool they provide to users for monitoring and quantifying the E/S/G aspect of a company’s performance.

They cover around 7,000 companies globally and combine over 250 ESG metrics with signals derived from over 50,000 sources across 170 countries in 4 languages. S-Ray tool applies a quantitative and algorithmic approach to ESG data which rates companies on the normative principles of the United Nations Global Compact (GC Score). They also provide an industry-specific assessment of companies' performance on financially material sustainability criteria (ESG Score). Both scores are combined with a "preferences" filter which can then be used to assess a company's business involvements.

- **GC Score** – based on the core principles of the UN Global Compact. This provides a deeper understanding of reputational risk facing a company
- **ESG Score** – industry specific analysis of a company’s performance on financially material E/S/G issues
- **Preferences Filter** – allows users to specify the activities they care about and check the business involvement of companies in these activities

**Figure 54. Arabesque S-Ray Process Flow**
**METHODOLOGY**

**Philosophy**

Arabesque’s mission is to take sustainable investing to the mainstream, pledging to make ESG data available to all. S-Ray allows investors to measure the sustainability of over 7,000 companies worldwide which enables stakeholders to make more informed decisions in building a more sustainable future.

**Sources of Data and Update Frequency**

Arabesque S-Ray collects a range of data from three types of sources:

- **Report-based metrics**
  - Over 250 reported metrics from non-financial disclosures, e.g. sustainability or integrated reports

- **News-based controversies**
  - Sustainability-related company news from over 50,000 public sources published in over 170 countries in four languages: English, French, Spanish and German.
  - All relevant news is organized by company and by topic, and then assigned a news value. This is a function of an article’s controversy level, how long ago it occurred, and the impact of the source

- **NGO-based activities**
  - NGO campaign activities across over 400 sustainability issues. NGO campaigns can be positive or negative in nature.

**How AI / Machine Learning is Applied**

There are three layers or steps in calculations of their ESG scores:

- **Input Layer**
  - Sourcing ESG data from third party data vendors

- **Feature Layer**
  - Data obtained from the Input Layer is aggregated into a set of ESG Feature Scores into a single ESG score systematically

- **Score Layer**
  - Combine features into Arabesque S-Ray scores

Machine learning is initially applied by Arabesque S-Ray’s news-based and NGO-based data providers, who use natural language processing algorithms to classify documents using keywords / bags of words techniques.

Next, machine learning is then applied in-house in the feature layer. The feature layer is introduced to further structure the input data along the below 22 sustainability topics (features) using semi-supervised dimensionality reduction techniques.
Typical dimensionality reduction techniques (e.g., PCA – principal component analysis) are classified as unsupervised machine learning techniques. Rather than applying these techniques unconditionally, Arabesque S-Ray’s algorithm requires human oversight where metrics are mapped to 22 features through applications of proprietary taxonomy, outcome and preparation are separated and only then they employ correlation / dimensionality reduction techniques. This way, it avoids spurious data aggregations such as combining E with S data in error.

Following on that, machine learning is applied in the approach used to determine which of these features are financially material to each company. Arabesque S-Ray then applies supervised machine learning to find which of the sustainability feature scores can help materially explain future stock price performance in addition to traditional financial factors.
ESG Criteria Used

As mentioned previously, after the Input Layer where Arabesque S-Ray sources ESG data from 3rd party vendors who might be using a variety of ESG criteria assessing companies, they further structure the sourced data into the 22 sustainability topics (features).

Scores

As depicted on the previous page, three layers are deployed before arriving at the final scores:

**Input Layer** – they collect a large number of data points (metrics) from 3rd party data providers. The data metrics are then grouped into four categories as below. Note that they do not produce any raw data metrics themselves.

---

**Figure 56. Arabesque S-Ray Groupings of Data Metrics**

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Report-Based</th>
<th>News-Based</th>
<th>NGO-Based</th>
<th>Materiality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio of female managers, CO₂ emissions, gambling revenue ... (see Appendix A for full list)</td>
<td>Long-term measure of company sustainability</td>
<td>Short-term adjustment (negative only)</td>
<td>Short-term adjustment</td>
<td>Materiality [i.e. Yes or No]</td>
</tr>
<tr>
<td>Example Data Sources</td>
<td>Company annual report / disclosure</td>
<td>For each article: (i) category, (ii) controversy score [1-3], (iii) impact</td>
<td>For each campaign: (i) category, (ii) sentiment [+2 to -2] (iii) response (iv) impact</td>
<td>SASB Materiality Map</td>
</tr>
<tr>
<td>Example Data Provider(s)</td>
<td>Reftinity</td>
<td>Media outlets (FT.com, reuters.com, forbes.com), ...</td>
<td>WWF, Amnesty International, Oxfam, Greenpeace, ...</td>
<td>SASB</td>
</tr>
<tr>
<td>How is data derived?</td>
<td>Report-based metrics are broadly quantitative in nature. Data is either delivered to Arabesque through an API/FTP from a 3rd party or (when available) Arabesque scrapes the data themselves from public reports. No discretion. Human involvement may be employed for quality checking, and/or input of data into databases)</td>
<td>Done entirely at the 3rd party level, articles are typically identified using natural language processing (e.g. algorithms which classify documents using key-word, or basket of word techniques). Human discretion may be employed at the third party level to oversee the severity and impact scores of the article (subject to defined guidelines).</td>
<td>NGO reports are typically identified using natural language processing (e.g. algorithms which classify documents using key-word, or basket of word techniques) under human supervision. Human discretion may be employed to quantify sentiment, impact and response of the NGO campaign (subject to defined guidelines).</td>
<td>S-Ray determines materiality of ESG factors using a supervised machine learning approach. As a starting point, factor weights are used from third-party materiality analyses that have been constructed using industry consultation (e.g. CEO panel), which includes human discretion. The further processing of these inputs is entirely rules-based with no further discretion at Arabesque.</td>
</tr>
<tr>
<td>Human Involvement?</td>
<td></td>
<td>Covalence</td>
<td>SIGWatch</td>
<td></td>
</tr>
<tr>
<td>Number of Data Points</td>
<td>250+ metrics</td>
<td>30,000+ news sources</td>
<td>60,000+ NGO campaigns</td>
<td>1500+ assessments</td>
</tr>
</tbody>
</table>

Source: Source: Arabesque S-Ray and their respective 3rd party data providers*

* Arabesque S-Ray has a preference not to publically disclose the full list of 3rd party data providers
Feature Layer – It is clear that the data collected from the Input Layer could have significant correlations and overlaps, which need to be addressed. Hence, the Feature Layer is introduced to further structure the input data along the 22 sustainability topics using the semi-supervised dimensionality reduction techniques highlighted earlier. Additional measures are introduced to ensure no dominant reliance on any one particular data vendor.

Feature Score
For each company which has sufficient data points, Arabesque S-Ray derives a Feature Score for each of the 22 Features mentioned earlier according to the following formula:

\[ F_{ik} = \overline{F}_{ik} \times (1 + CF_{ik}^{\text{News}} + CF_{ik}^{\text{NGO}}) \]

where

\[ F_{ik} \] = The Feature score with respect to company i and Feature k, expressed as a number between 0 and 100.

\[ \overline{F}_{ik} \] = The Report-based Feature Score with respect to company i and Feature k, expressed as a number between 0 and 100.

\[ CF_{ik}^{\text{News}} \] = The News-Based Correction Factor (%) with respect to company i and Feature k, which ranges in value between 0% and -50%.

\[ CF_{ik}^{\text{NGO}} \] = The NGO-Based Correction Factor (%) with respect to company i and Factor k, which ranges in value between +25% and -25%.

The starting point for the Feature Score construction is report-based metrics. These metrics are generally updated relatively infrequently (e.g. annually). In an attempt to make the Feature Score more responsive, the report-based feature score is adjusted using two correction factors which are calculated daily and derived from news-based and NGO-based data sources, looking back one year. The news-based Correction Factor and NGO-Based Correction Factor are in percentages and updated daily which adjust the report-based Feature Score based on recent news flows and NGO campaigns that are relevant to a given company. The report-based Feature Score and two correction factors are constructed as follows:

- **Report-based Feature Score**

The report-based Feature Score is a value between 0 and 100, which is constructed from approximately 250 individual report-based metrics. Each of the report-based metrics has a static mapping to an ESG Feature. For example, the table below lists the report-based metrics that are mapped to the Emissions Feature.

### Figure 57. Arabesque S-Ray Emissions Topics

<table>
<thead>
<tr>
<th>Report-Based Metric</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO2 Emissions</td>
<td>Emissions</td>
</tr>
<tr>
<td>CO2 Reduction Initiatives</td>
<td>Emissions</td>
</tr>
<tr>
<td>Consideration of Climate Change Risks and Opportunities</td>
<td>Emissions</td>
</tr>
<tr>
<td>Environmental Impact of Production Processes</td>
<td>Emissions</td>
</tr>
<tr>
<td>F-Gases Emission Reduction Initiatives</td>
<td>Emissions</td>
</tr>
<tr>
<td>NOx and Sox Emissions Reduction Initiatives</td>
<td>Emissions</td>
</tr>
<tr>
<td>Ozone-Depleting Substances Reduction Initiatives</td>
<td>Emissions</td>
</tr>
<tr>
<td>VOC Emissions Reduction Initiatives</td>
<td>Emissions</td>
</tr>
<tr>
<td>Air Emissions Reduction Programs</td>
<td>Emissions</td>
</tr>
<tr>
<td>Company Transportation Emissions Reduction Program</td>
<td>Emissions</td>
</tr>
<tr>
<td>Emissions Objectives</td>
<td>Emissions</td>
</tr>
<tr>
<td>Emissions Performance Monitoring</td>
<td>Emissions</td>
</tr>
<tr>
<td>Emissions Reduction Policy</td>
<td>Emissions</td>
</tr>
</tbody>
</table>

Source: Arabesque S-Ray
Next, for each stock and feature, the ESG Feature score is constructed by aggregating all the metrics for the given feature into a normalized score between 0-100 (typically using a ranking-based approach). A distinction is made between preparation-based (e.g. an emissions reduction policy) and outcome-based metrics (e.g. scope 1 emissions). Dimensionality or correlation amongst inputs is also taken into account to avoid overweighting.

- **News-based Correction Factor**

On a daily basis, Arabesque S-Ray consumes data from third party data providers who identify and classify news articles. For each article that is flagged as having ESG-related content, their providers label the sustainability topics which are discussed in the article, issue a measure of the severity of the controversy (ranging from 1 to 3), and then provide a measure of the articles impact (ranging from 1 to 3) of the source.

S-Ray uses this information to calculate a News Value (NV) score for each relevant article observed in the past one year. For this factor, only negative news articles are included. The News Value is then adjusted by a decay function reflecting the age of the article as well as the current level of the long-term (Report-based) score. The output of this step is the Present News Value (PNV). S-Ray identifies the article with the largest PNV and normalizes it to a correction factor between -50% to 0%.

- **NGO-based Correction Factor**

The construction of the NGO-based Correction Factor is very similar to that of the News-based Correction Factor. Arabesque S-Ray ingests data from third party data providers who monitor articles related to NGO-campaigns.

For each article that is flagged as having ESG-related content, providers label the sustainability topics which are discussed in the article and include a measure of the severity (positive or negative) of the campaign, whether or not the targeted company is partnering on the NGO campaign and finally the impact of the NGO(s). Arabesque S-Ray uses this information to calculate a Campaign Value (CV) score for each relevant article. The Campaign Value is then adjusted by a decay function reflecting the age of the article as well as the current level of the long-term (report-based) score. The output of this step is the Present Campaign Value (PCV). S-Ray identifies the article with the largest PCV and normalizes it to a correction factor between -25% to +25%.

These short-term correction factors are needed to address the slowness of the report-based long-term trend. The final feature scores are then the 22 long-term trend scores (ranging from 0 to 100) multiplied by their respective short-term correction factors which consist of the news-based controversies and NGO campaign activities.

Note that as not all companies are covered by the third party data providers sourced by Arabesque S-Ray, in order for a company to have a score in S-Ray, there are several requirements as below:

- Depending on the feature, either at least one outcome- and/or preparation-based input would be needed
- For a report-based feature score, at least two inputs are required
- For a news / NGO-based feature score, only one input is required
- Total feature scores can only be calculated when a report-based feature is present, i.e. companies only getting news / NGO data will not get a feature score

**Score Layer** – Arabesque S-Ray produces three scores to address different ESG needs from the users: GC Score, ESG Score and Preferences Filter. The minimum requirements for producing scores are:

- At least one feature score available to calculate a sub-score
- Total ESG and GC scores would only be computed for companies that have coverage on all sub-scores
The Rise of AI in ESG Evaluation | 47

**GC Score**

The GC score provides a normative assessment of companies based on the four core categories of the ten principles of the United Nations Global Compact (GC): human rights, labour, environment and anti-corruption.

In order to calculate the GC score, they firstly map the relevant features into each of the four GC categories. A distinction is made to separate the features that focus on negative aspects from those more positive in nature, with the former taking precedence in case of poor performance. For example, if a firm is engaging in human rights violating activities but at the same time donating money through its foundation, Arabesque S-Ray's algorithm would almost completely discard the positive features and emphasise the more negative ones. The resultant GC scores (ranging from 0 to 100) quantify a company's performance in each of these four GC categories.

In addition to the individual four GC scores, Arabesque S-Ray also provides an aggregated GC score using a non-compensatory approach which reflects the nature of the GC principles. Each category starts off with a 25% weight but gets more weight allocated as the score drops below 50 which is deemed to be the neutral centre. Consequently, it is not possible to compensate poor performance in one category with good performance in another. As a company's performance deteriorates in any of the four categories, more weights are assigned to that category which would then drive the overall GC score.

For all GC scores, they are scaled between 0 and 100 with higher numbers indicating better behaviour/performance and vice versa. These scores can then be used to identify outperformers vs those who are exposed to reputational risk on the downside.

**ESG Score**

The ESG Score for a given company ranges from 0 to 100, which is constructed as the weighted sum of Feature Scores and Materiality Weights.

\[
ESG_i = \sum_{k=1}^{22} \text{weight}_{i,k} \times F_{i,k}
\]

Where

\[
\text{Weight}_{i,k} = \text{Materiality Weight, expressed as a percentage, with respect to Feature } k \text{ and company } i. \text{ The sum of weights for a given company equals to 100%}.
\]

\[
F_{i,k} = \text{Feature Score with respect to Feature } k \text{ and company } i.
\]

The UN SDGs are not explicitly used in constructing their UNGC and ESG scores but Arabesque S-Ray is currently developing an SDG Quantification Tool for their users. SASB criteria are used in their materiality assessment as a starting point and then the model is calibrated on a quarterly basis to capture changing materiality and monitor data quality over time.

In the aforementioned equation, we have discussed the feature scores in the Feature Layer section and now we turn our attention to the construction of materiality weight. Materiality weight is an important determinant to S-Ray's ESG Score. In essence, it is aimed to identify whether the nature of the data input is significant, relevant and material with respect to a given company, industry or sector. For example, CO2 emissions are an important ESG factor in the Industrials and Energy sectors but less relevant for IT and Financials.

On a quarterly basis, S-Ray updates the weight in regard to a given feature \(k\) for a given company \(i\).

\[
\text{weight}_{i,k} = \text{baselineWeight}_{i,k} + \text{dynamicWeight}_{i,k}
\]

where

\[
\text{baselineWeight}_{i,k} = \text{The Baseline Materiality Score with respect to Feature } k. \text{ A Boolean which reflects the materiality of a Feature, based upon assessments of third-party sources (e.g. SASB Materiality Map and 3rd party industry research reports) with respect to the industry and sector of company } i.
\]
The Dynamic Materiality Score with respect to Feature k and with respect to company i. The Dynamic Materiality Score, which is updated quarterly, seeks to capture how the materiality of a Feature varies through time. The Dynamic Materiality Score is a measure of a Feature’s statistical importance at explaining historical sector and industry group price performance (beyond what can be explained by a traditional 3-factor financial model).

The Baseline Materiality Score is assigned based on 3rd party research and industry reports on materiality. One valuable source is the SASB Materiality Map, which is constructed by ranking issues in each industry based on two types of evidence (1) investors in the industry are interested in the issue and (2) the issue has the ability to impact companies within the industry. This serves as a good starting point to determine which categories are material in the companies’ respective industries.

The calculation of the Dynamic Materiality Score is based upon the Fama-French 3-factor asset pricing model, which describes historical returns as a linear combination of returns from three factors: (1) market returns, (2) size (small – large), and (3) value (high – low):

\[
   r_A = R_f + \beta_m(R_m - R_f) + \beta_{size}R_{size} + \beta_{value}R_{value} + \epsilon_A
\]

where

- \( R_f \) = Riskfree Rate
- \( R_m \) = Return of Market
- \( R_{size} \) = Excess Return of Size factor (Small Cap stocks over Large Cap stocks)
- \( R_{value} \) = Excess Return of Value factor (High Book-to-Price stocks over low Book-to-Price)
- \( \beta_m \) = Beta to Market factor
- \( \beta_{size} \) = Beta to Size factor
- \( \beta_{value} \) = Beta to Value factor
- \( r_A \) = Return with respect to the timeseries A
- \( \epsilon_A \) = Residual return with respect to the timeseries A

For a given company i, S-ray identifies 12 historical time series corresponding to the company’s sector and industry group (based on Factset industry classifications), over the prior 1-year, 3-year and 5-year periods, and where stocks within each sector or industry group are weighted on both a cap-weighted basis and an equal-weighted basis. For each of the 12 time series, the above 3-factor model is calibrated using a regression model. The outputs of this step are the beta parameters (\( \beta_m, \beta_{size}, \beta_{value} \)) and a time series of residual returns (\( \epsilon_A \)). These residual returns represent idiosyncratic sector or industry returns which cannot be explained by the three Fama-French factors.

S-Ray then identifies which ESG Features are the most effective at explaining the residual returns of the relevant sectors & industry groups. This is achieved by first constructing long-short stock portfolios for each Feature, e.g. a portfolio which goes long companies with the best diversity scores, and short those with the worst. The performance of each of these long/short portfolios is then tested against the residual time series to derive materiality. A recursive feature elimination approach with cross-validation (RFECV) is performed for each of the 12 historical time series of residual returns. The result of each time series is a binary vector highlighting which features are material (assigned 1) and which are not (assigned 0). The optimal number of features to include is determined by a cross-validation approach (i.e. testing the best threshold value). The vectors from each time series are then averaged to derive the Dynamic Materiality Score for each Feature.

The baseline weight is then combined with the dynamic weight and normalised to arrive at the total feature weight, which is used in calculating the final ESG score – a weighted sum of the feature scores multiplied by respective total feature weights.

S-Ray also provides the E/S/G sub-scores which are calculated by considering only the features within each of these pillars. As per the GC score, the E/S/G sub-scores and ESG score are scaled between 0 and 100 with the higher scores indicating better company performance.
In summary, their GC Score take a normative approach to sustainability and represents reputational risk of companies, while the ESG score is calibrated based on financial materiality. From an investor’s perspective, the former is better suited for downside protection and the latter could be used to identify both outperformers and underperformers.

Preferences Filter

The last output from the Score Layer is Preferences Filter which covers a set of 11 business involvements (see below) for users to check companies against these activities. Based on companies’ revenues, Arabesque S-Ray transforms them into 11 involvement flags which are binary in nature – 1: yes, 0: no. These flags can be used to exclude companies which participate in these activities, well-suited for asset managers who wish to use binary negative-screening/exclusions based on ESG in their portfolios.

### Figure 58. Arabesque S-Ray 11 Business Involvements for Preferences Filter

<table>
<thead>
<tr>
<th>Business Involvement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult entertainment</td>
<td>Does the company derive significant revenues from adult entertainment products?</td>
</tr>
<tr>
<td>Alcohol</td>
<td>Does the company derive significant revenues from the production and/or sale of alcohol?</td>
</tr>
<tr>
<td>Defence</td>
<td>Does the company derive significant revenues from defence contracting?</td>
</tr>
<tr>
<td>Fossil Fuel</td>
<td>Does the company derive significant revenue from the extraction of fossil fuels?</td>
</tr>
<tr>
<td>Gambling</td>
<td>Does the company derive significant revenues from gambling products and/or services?</td>
</tr>
<tr>
<td>Gmo</td>
<td>Does the company significantly engage in research and/or production of genetically modified organisms (GMO) based products?</td>
</tr>
<tr>
<td>Nuclear</td>
<td>Does the company derive significant revenue from non-military uranium enrichment and/or the exploitation of nuclear energy?</td>
</tr>
<tr>
<td>Pork</td>
<td>Does the company derive significant revenues from the sale and/or production of pork-based products?</td>
</tr>
<tr>
<td>Stem Cells</td>
<td>Does the company derive significant revenues from stem cell research?</td>
</tr>
<tr>
<td>Tobacco</td>
<td>Does the company derive significant revenues from the sale and/or production of tobacco products?</td>
</tr>
<tr>
<td>Weapons</td>
<td>Does the company significantly engage in the sale and/or production of weapons?</td>
</tr>
</tbody>
</table>

Source: Arabesque S-Ray

Company Examples

The two examples below demonstrate how Arabesque S-Ray’s scores pick up the deteriorations of corporate behaviour on the ESG front. The first example is Nissan, where their GC score, Anti-Corruption sub-score and Governance scores all dropped significantly before July 2018 when Nissan admitted they altered the cars’ emissions and fuel efficiency data. Later that year, its chairman was arrested amid accusations of under-reporting his compensation. Both events affected the share price substantially, resulting in 4.6% and 5.5% drops respectively.

### Figure 59. Arabesque S-Ray ESG Score Example - Nissan

Source: Arabesque S-Ray
The second example is SunEdison, which filed for bankruptcy in April 2016. SunEdison was a renewable energy company and attracted a high Environment score. However, during the 5 years prior to its bankruptcy, Arabesque S-Ray's Governance score already showed a gradual deterioration relative to its sector dropping from the 88th percentile to the 31st percentile until it reached 0 at the beginning of 2016. In March 2016, it was reported that investigations were launched into SunEdison's cash position which was suspected to have been overstated. In addition, following multiple acquisitions over the years, the company was loaded with more than US$ 11 billion in debt, which also contributed to its downfall. The share price dropped from over US$33 in July 2015 to $0.34 in April 2018.

Figure 60. Arabesque S-Ray Example - SunEdison

![Graph showing SunEdison's ESG scores]

- Despite a high Environment ESG sub-score, in 2015, SunEdison’s Governance ESG sub-score sector percentile dropped from 88% to 31% and continued falling until it hit 0% in early 2016.
- In March, reports highlighted investigations regarding SunEdison overstating its cash position.
- SunEdison attempted to acquire Vivint Solar for $2.2 billion, a 52% premium, which flagged investors of SunEdison’s growing debt alongside its increasing number of acquisitions.
- By the end of March, the stock price went from over $30/share (2015) to $0.36/share.
- In April, the stock was delisted from NYSE and the company filed for bankruptcy.

Source: Arabesque S-Ray
**EMPIRICAL ANALYSIS**

**Coverage**

As noted in the previous section, Arabesque S-Ray collects 250+ metrics from other ESG providers, combined with large numbers of other news-based and NGO-based sources. Consequently, we expect their data coverage of the two universes to be broad as they aggregate across data sources they have through the mechanism described earlier.

Judging from the charts in Figure 61 and Figure 62, they indeed appear to confirm that's the case. For developed world, Arabesque S-Ray has achieved over 90% of coverage since 2010 and even with emerging markets, the coverage had been consistently above 80% but dropped in late 2018 into 2019 as MSCI started to include China A Shares into their EM benchmark.

![Figure 61. Arabesque S-Ray Coverage: World](source: Arabesque S-Ray, CGDI, MSCI)

![Figure 62. Arabesque S-Ray Stock Coverage for EM](source: Arabesque S-Ray, CGDI, MSCI)

As before, we also check the coverage by market cap bucket to examine if there are any pronounced size biases in their data. The charts below show that for developed markets, the distribution of their coverage is pretty evenly spread across various market cap buckets, while there is an exception to Emerging Markets for companies under US$750M largely due to the China A shares inclusion.

![Figure 63. Arabesque S-Ray Coverage by Market Cap Bucket](source: Arabesque S-Ray, CGDI, MSCI)

![Figure 64. Arabesque S-Ray Stock Coverage by Market Cap Bucket](source: Arabesque S-Ray, CGDI, MSCI)

Checking the coverage from the sector perspective over time, the tables below indicate very balanced coverage across all sectors in developed markets with majority of them achieving more than 80% of coverage regardless of by market-cap weighting or equal weighting. For emerging markets, the data coverage is also very good, with the only exception being late 2018 & 2019 when China A shares were included in the index.
In terms of the tilts embedded in Arabesque S-Ray ESG scores, we examine the score distribution of ESG score and individual E/S/G scores across sectors. Figure 67 to Figure 70 show the analysis results from the developed markets. Generally speaking the overall ESG mean scores appear to centre around the 50 – 55 range, with very similar standard deviations across sectors. Zooming into the E/S/G scores individually, the variations then become much more apparent. For E scores, the highest average scores are observed in Materials and Utilities sectors which indicate the environmental issues their business activities need to address is indeed the core focus in these sectors and thus are reflected in their scores. The biggest variations in scores are seen in Information Technology and Consumer Discretionary sectors which could be due to the broad nature of these sectors. Compared to E scores, the S scores have substantially smaller standard deviations with Materials and Utilities together with Consumer Staples having the highest average scores. With regard to G scores, the range of standard deviations is between those of E and S scores with very consistent mean scores around 50. The biggest variation is within Consumer Discretionary sector, followed by Industrials and Financials, reflecting the diverse nature within these sectors where corporate governance could vary greatly in the sub-industries.

Source: Arabesque S-Ray, CGDI, MSCI

Figure 67 to Figure 70 show the analysis results from the developed markets. Generally speaking the overall ESG mean scores appear to centre around the 50 – 55 range, with very similar standard deviations across sectors. Zooming into the E/S/G scores individually, the variations then become much more apparent. For E scores, the highest average scores are observed in Materials and Utilities sectors which indicate the environmental issues their business activities need to address is indeed the core focus in these sectors and thus are reflected in their scores. The biggest variations in scores are seen in Information Technology and Consumer Discretionary sectors which could be due to the broad nature of these sectors. Compared to E scores, the S scores have substantially smaller standard deviations with Materials and Utilities together with Consumer Staples having the highest average scores. With regard to G scores, the range of standard deviations is between those of E and S scores with very consistent mean scores around 50. The biggest variation is within Consumer Discretionary sector, followed by Industrials and Financials, reflecting the diverse nature within these sectors where corporate governance could vary greatly in the sub-industries.
For emerging markets, the observations are not that dissimilar to those of the developed markets in terms of the mean score distributions. However, the variation of scores tend to be larger in EM especially the G scores which might indicate the standard of corporate governance is not as mature as that in developed markets and the lack of transparency and business ethics are still common issues in those markets.
Figure 71. Arabesque S-Ray Overall ESG Score Distribution (bar shows +/- 1 std, circle denotes mean) - EM

Figure 72. Arabesque S-Ray E Score Distribution (bar shows +/- 1 std, circle denotes mean) - EM

Figure 73. Arabesque S-Ray S Score Distribution (bar shows +/- 1 std, circle denotes mean) - EM

Figure 74. Arabesque S-Ray G Score Distribution (bar shows +/- 1 std, circle denotes mean) - EM

Source: Arabesque S-Ray, CGDI, MSCI

Source: Arabesque S-Ray, CGDI, MSCI
Performance

We have established that Arabesque S-Ray has decent coverage of World and Emerging Markets universes, thanks to their sourcing of data across many ESG third party providers and other sources. With this in mind, we examine whether or not the comprehensive coverage translates into outperformance using their overall ESG scores. To that end, we construct quintile portfolios as per the analysis of two previous vendors based on the overall ESG scores and also the individual E/S/G scores to measure the efficacy of such scores linking with stock performance.

Figure 75 and Figure 76 show the quintile performance and Quintile 5 vs Quintile 1 relative to MSCI World indices. The quintile performance on the left-hand chart depicts monotonic increasing returns from Quintile 1 to Quintile 5, even though there is little difference amongst quintiles 2 – 4. Nonetheless, the positive performance spread from (Quintile 5 – Quintile 1) has been gradually increasing since 2015. Compared to MSCI World ESG Leaders and World indices (both market cap-weighted and equal-weighted), it can be seen that the extent that Quintile 5 outperforms the indices is moderately greater than the underperformance from Quintile 1 vs the benchmarks. On a risk-adjusted basis, Quintile 5 has the largest Sharpe Ratio while Quintile 1 has the lowest with similar volatility levels as the benchmarks’. This suggests that Arabesque S-Ray’s ESG scores are indeed suited for both choosing best-in-class companies and identifying poor ESG performing companies.

<table>
<thead>
<tr>
<th>Figure 75. Arabesque S-Ray ESG Score Quintile Performance</th>
<th>Figure 76. Arabesque S-Ray ESG Top/Bottom Quintile Performance vs World benchmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Quintile Performance Chart" /></td>
<td><img src="image2" alt="Quintile Performance vs World Benchmarks Chart" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Figure 77. Arabesque S-Ray Strategy Performance Statistics - World</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Annualised Returns</strong></td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>9%</td>
</tr>
<tr>
<td><strong>Annualised Volatility</strong></td>
</tr>
<tr>
<td><strong>Sharpe Ratio</strong></td>
</tr>
</tbody>
</table>

Source: Arabesque S-Ray, CGDI, MSCI

The results unfortunately are not as encouraging when we construct the same quintile strategies in emerging markets. Despite having good coverage of the universe, the quintile performance is not monotonic even though Quintile 5 does beat Quintile 1, but it is Quintile 2 that underperforms the most. When compared to MSCI benchmarks, both Quintile 5 and 1 outperform the broad emerging market indices but just like strategies based on data from the two previous vendors, they fail to beat MSCI EM ESG Leaders index. Once again, this highlights potential issues of lack of frequent newsfeeds and other data sources that are ESG-centric, reflecting perhaps a lower degree of ESG awareness in these markets.
Next we examine the Long-Short performance based on the individual E, S, and G scores as Arabesque S-Ray provides such breakdown on a comparative basis across companies within the same sector. Figure 80 shows that for developed world, out of E/S/G, G scores are the most effective measure in predicting stock performance. Both E and S have not performed during the testing period we have but if excluding the period prior to 2012, the performance of long-short strategies constructed based on E and S scores has been flat.

The Long-Short performance of E/S/G in emerging markets reflects the modest return profile witnessed earlier but once again the performance breakdown shows that G is the most effective factor in delivering positive performance as a strategy, even though the magnitude is rather modest. Interestingly, both E and S strategy performance improved significantly since 2015 coinciding with the inflection point of ‘ESG Investing’ we highlighted in the introduction section of the study. This can be seen as additional evidence that the attention from investors in EM also shifted towards more environmental and social awareness around corporate behaviour.

To understand where the outperformance is coming from for both the overall ESG score and also Governance score on its own, as before we employ the Fama-French-Carhart regression to dissect the outperformance into market beta, value, size and momentum exposures and examine if there is additional alpha to be had based on these scores after taking the factor tilts into account.

The table overleaf shows for the overall ESG score, the only statistically significant tilt at 95% is the large cap bias. While the other factors are not significant, after accounting for their contributions to the strategy performance, the residual return is not statistically significant at 90% confidence level. For G score, it is interesting to see how different the tilts are compared to those of the overall ESG.
score. The return profile of the long-short strategy based on G score has tilts towards lower beta, smaller cap and higher price momentum. The residual return is statistically significant at 90%, indicating there is additional alpha embedded in their G score. This finding can guide asset managers to focus on the governance aspect of companies as good governance appears to attract higher returns beyond the common risk premia.

Broadening to include other common style factors, Figure 83 shows the highest correlations for both the overall ESG and G scores are with Quality factor, i.e. higher ESG/G scores capture high quality companies, which is intuitive. The return of the long-short strategy based on the Governance score is highly correlated with high price momentum and lower risk as well.

Figure 82. Fama-French-Carhart Regression on ESG and G Score Returns – World

<table>
<thead>
<tr>
<th>ESG</th>
<th>Coefficients</th>
<th>Standard Error</th>
<th>P-value</th>
<th>G</th>
<th>Coefficients</th>
<th>Standard Error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.002</td>
<td>0.001</td>
<td>0.208</td>
<td>Intercept</td>
<td>0.003</td>
<td>0.001</td>
<td>0.060*</td>
</tr>
<tr>
<td>MKT</td>
<td>0.054</td>
<td>0.034</td>
<td>0.118</td>
<td>MKT</td>
<td>-0.082</td>
<td>0.037</td>
<td>0.031**</td>
</tr>
<tr>
<td>Value</td>
<td>0.108</td>
<td>0.099</td>
<td>0.282</td>
<td>Value</td>
<td>-0.061</td>
<td>0.109</td>
<td>0.576</td>
</tr>
<tr>
<td>Size</td>
<td>0.381</td>
<td>0.165</td>
<td>0.023**</td>
<td>Size</td>
<td>-0.494</td>
<td>0.181</td>
<td>0.007*</td>
</tr>
<tr>
<td>PMOM</td>
<td>0.118</td>
<td>0.088</td>
<td>0.180</td>
<td>PMOM</td>
<td>0.324</td>
<td>0.096</td>
<td>0.001**</td>
</tr>
</tbody>
</table>

* Significant at 90%
** Significant at 95%

Figure 83. ESG & G Score Return Correlations with Risk Premia

<table>
<thead>
<tr>
<th>ESG</th>
<th>G</th>
<th>Value</th>
<th>Size</th>
<th>PMOM</th>
<th>Quality</th>
<th>Low Risk</th>
<th>EMOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.15</td>
<td>-0.23</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>0.24</td>
<td>-0.18</td>
<td>-0.27</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMOM</td>
<td>0.06</td>
<td>0.42</td>
<td>0.10</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality</td>
<td>0.27</td>
<td>0.51</td>
<td>0.02</td>
<td>0.47</td>
<td>0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Risk</td>
<td>-0.08</td>
<td>0.42</td>
<td>-0.02</td>
<td>0.03</td>
<td>-0.13</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>EMOM</td>
<td>-0.15</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.05</td>
<td>-0.11</td>
<td>0.01</td>
<td>-0.51</td>
</tr>
<tr>
<td>Growth</td>
<td>0.10</td>
<td>-0.13</td>
<td>0.36</td>
<td>-0.11</td>
<td>0.01</td>
<td>-0.51</td>
<td>-0.14</td>
</tr>
</tbody>
</table>

Source: Arabesque S-Ray, CGDI, MSCI

9 Risk premia returns are based on pure style index returns constructed by Citi Quant Research team
SUMMARY

In this section, we have explored the data sources, methodology and constructions of Arabesque S-Ray's ESG offerings, followed by empirical analysis on their various scores. Compared to the other vendors, Arabesque S-Ray does not have its own metrics that are derived from the raw data sources, but instead it aggregates the output from many ESG third party providers across report-based, news-based and NGO-based sources and derives the overall ESG scores and sub-components by deploying AI / machine learning techniques on top of these datasets.

Thanks to its aggregation methodology, Arabesque S-Ray is able to achieve high percentages of universe coverage for both MSCI World and Emerging Markets universes. Dissecting the data coverage across sector and market cap spectrums, there is no evidence of pronounced sector tilts or size tilts apart from the lowest market cap bucket due to the inclusion of China A shares and limitations of data sources for those stocks.

Examining the average sector scores and standard deviations of their ESG and individual E/S/G scores, the results show both similarities and divergence between developed markets and emerging markets. They are similar in terms of the range of sector mean scores and different in terms of variation of scores in their E/S/G components. Specifically, the Environmental scores have much larger standard deviations in developed markets on a sector level while the Governance scores claim the top spot in EM in terms of variations amongst sectors. This could indicate environmental issues are under much more scrutiny in developed markets while the same is true for governance issues in EM, where business ethics and transparency is lacking on a relative basis.

We have also tested the efficacy of Arabesque S-Ray's overall ESG and sub-component scores as signals to construct strategies that would outperform the MSCI benchmarks. The results for DM are decent in that the using their overall ESG scores to construct quintiles does yield the desired monotonic increasing return profile from the least attractive to the most attractive quintile. Relative to the chosen MSCI benchmarks, the top quintile has beaten all of indices since 2015 while the bottom quintile has underperformed. The analysis in EM is not as encouraging despite decent coverage of the universe. While the top quintile does indeed outperform the bottom quintile, the return profile is far from monotonic, and the top quintile fails to beat MSCI EM ESG Leaders index. Using the individual E, S, G scores, we show that long-short strategies based on G scores achieve the highest returns in both developed markets and EM, suggesting that the markets pay the most attention to governance issues and reward companies with good governance accordingly.

While our analysis demonstrates Arabesque S-Ray scores can be used to add alphas to investment strategies, there are several caveats investors should bear in mind:

- As the underlying data used to construct Arabesque S-Ray overall ESG score is sourced from third party providers, the process of how these third party providers generates their scores is not transparent.
- Even though they have set minimum data requirements for a company to have a score to ensure continuation of coverage, the underlying make-up of the third party providers might increase or decrease over time, which would make the continuity of ESG scores based on the same sources more challenging.
- The information gathering process by third party data providers may fail to accurately identify and categorise data, or the methodology applied could be stale.
**Analyst-driven vs AI-led**

As set out in "Motivation and Scope", this paper is intended to shed some light in terms of how AI-led ESG providers could potentially address the shortcomings of analyst-driven, traditional ESG vendors. Armed with the knowledge we have acquired in the previous sections on the three individual vendors, we examine below the issues associated with analyst-driven vendors and critique whether or not the AI vendors resolve or at least mitigate them.

- **Lack of consistency and comparability in companies’ disclosed information**
  
  We highlighted earlier, inconsistency in company reporting also led to inconsistent interpretation of the analyst-driven ESG ratings as such models would be prone to subjectivities. With AI-led models, all three vendors have put forward their own proprietary methods of mapping companies’ behaviour onto their ESG issue/topics from the outset and apply systematically to score companies using AI/machine learning techniques. Compared to the analyst-driven models, this would ensure the consistency is applied to the companies these vendors cover. Having said that, the fact that each of them has different ways of assessing companies might still not be ideal if investors are looking for a gold standard to evaluate corporates ESG behaviour, just like there are common reporting standards (US GAAP, IFRS etc.) for financial information of companies. What we do observe from our discussions with these AI vendors is that there appears to be a convergence towards SASB and UN SDGs as all of them are either using these criteria already or working towards fully integrating them into their scoring processes.

- **There are no common reporting standards mandating companies to disclose**
  
  Clearly, this is not an issue which can be resolved by AI-led ESG data providers per se but their usage of big data combined with AI / machine learning techniques makes it possible to at least assess companies in a systematic way which is defined at the outset of their ESG scores construction. By utilising large volumes of data sources, their approaches are highly scalable and consistent across the companies they cover, which could provide valuable insights that might not have been possible if solely relying on disclosures made by companies, as such disclosures are unaudited and susceptible to manipulations.

- **Hard to measure companies’ intra-year ESG performance and to take actions based on annual disclosures**
  
  This is one key advantage of using AI-led data providers as they employ external data sources such as newsfeeds, blogs, NGO reports etc to evaluate companies from an outside-in perspective. The deployment of these sources makes it possible to provide near real-time assessment on corporate behaviour, which could not be achieved by using disclosure-based information from companies, which are typically available on an annual basis.

- **Human analysis makes assessments less timely and volume limiting**
  
  The major advantage of AI-led models is that they are highly scalable and efficient when expanding stock universes. Having said that, this does not mean there is no human input into the process as ESG valuation does require significant domain expertise, and the AI vendors acknowledge that. Many of them apply the domain knowledge from their analysts/researchers at the beginning of the scoring process, working with their data engineers to fine-tune the algorithms in turning unstructured data sources into useful insightful ESG scores and ratings.
Lack of transparency in how ESG index providers derive their scores

One of the main criticisms of analyst-driven traditional ESG data providers is the lack of transparency in how they incorporate disclosure-based information and other sources to produce their own scores and ratings. Despite this major shortcoming, according to MSCI they provide their ESG ratings to over 90% of top 50 global asset managers. Also, with the main source of information being companies’ disclosures, it is puzzling for asset managers as to why the ratings could differ substantially across providers. For example, CSRHub has found that the correlation between MSCI’s and Sustainalytics’s ESG ratings is merely 0.32 for companies in S&P Global 1200 index.

AI-led ESG vendors on the other hand have quite similar data sources and methodologies, which one would expect should lead to higher correlations amongst them. However, our findings are quite the opposite (see Figure 84) where we observe lower rank correlations for the vendors highlighted in this report. This might at first appear to be surprising but we believe the reason for the low correlation is due to the fact that their scores are designed for different ESG applications. For example, RepRisk tasks themselves with flagging and alerting investors on risk incidents, i.e. the downside, while Truvalue and Arabesque S-Ray aim to provide scores on companies from both positive and negative aspects. With the latter two, there is a stark difference in their score constructions where Truvalue offers three ESG scores that suit different investment horizons and purposes while Arabesque S-Ray starts with report-based metrics, aggregates a large number of ESG data providers and then apply their proprietary algorithms to arrive at their overall ESG scores.

Figure 84. AI Vendors’ ESG Scores Rank Correlations

This finding might be regarded as a disappointment to asset managers who would like to see coherent ESG ratings and scores across vendors. While the current status quo with the analyst-driven providers does suggest depending on which vendor one utilise, companies could be rated dramatically differently. With the AI-led vendors, we see them as powerful ways to augment and complement the analyst-driven vendors, as they provide real-time and unbiased outside-in assessments that suit different ESG objectives and investment horizons. The lack of correlations amongst the three AI vendors highlighted in this report could in fact be useful from a diversification perspective.

In Figure 85, we show the strategy performance based on the three AI vendors and the indices which have been constructed utilising analyst-driven ESG ratings and scores\textsuperscript{10}. The chart below demonstrates the potential value-add these AI vendors could bring to investment performance, as higher frequency in updating the scores based on recent events could give them an edge over slower and more subjective analyst-driven scores.

\textsuperscript{10} Sustainalytics Leaders portfolio is constructed by Citi Quant Research team who utilise the same methodology as per MSCI’s Leaders index.
Having said that, there are issues with AI-led ESG vendors in that their reliance on newsfeeds, media and other sources means the lack of data or languages supported by these firms could significantly limit their coverage and quality of scores. The lack of data could be down to two possibilities: lack of attention in local markets on ESG issues, and smaller companies are reported much less frequently than larger companies. On the language issue, this is particularly important when assessing companies in emerging markets if the local language is not supported by the vendor. We believe these issues contribute to the less than satisfactory performance in EM we saw earlier in the report regardless of which vendor’s data we use. In our view, these observations show that it makes sense to combine both analyst-driven data with AI-led scores as the latter could correct the biases the former introduces and provides more up-to-date pictures of companies’ ESG performance. At the same time, analyst-driven data could bridge the gap for smaller companies where AI-led providers might not have an information advantage, due to the lack of data coverage.

Another common criticism of the analyst-driven ESG scores is that larger companies tend to have higher ESG scores. The justifications put forward by analyst-driven ESG providers are that large companies are typically in better shape financially and have the means to invest in ESG-related activities, hence improve their corporate profiles. This puts smaller caps at a distinct disadvantage as their resources are more limited and they might not be able invest as heavily as their larger counterparts. The AI ESG vendors we have highlighted in the report all recognise the potential size biases and adjust their scoring mechanisms accordingly to mitigate such biases. In Figure 86, for example, we show the mean scores by market cap bucket for the developed world universe based on RepRisk’s current RRI, Truvalue’s Insight score and Arabesque S-Ray’s overall ESG score. Note that RepRisk’s current RRIs typically cluster around 25, denoting low to medium risk. As can be seen below, Truvalue shows very little variation across market cap buckets. RepRisk actually has higher risk scores for larger caps, while Arabesque S-Ray exhibits a slight upward tilt towards larger caps, which might be due to their inclusion of report-based ESG providers. Nevertheless, compared to analyst-driven ESG scores, the size bias is substantially less pronounced in AI-led ESG scores.
The ultimate solution to the inconsistency in company reporting and in the ESG scores and ratings provided by different vendors would be to instigate compulsory reporting standards in ESG, which may still take some years to materialise. However, in the interim, we believe the combination of analyst-driven and AI-led models, which enables investors to have the flexibility to shift the emphasis of ESG scores from one model to the other or vice versa, could serve as the “best of both worlds” to suit their investment horizons and processes.

Figure 86. Average Scores by Market Cap Bucket – World

Source: RepRisk, Truvalue, Arabesque S-Ray, CGDI
Conclusion

Since 2015 ESG investing has attracted a lot of attention from investors, coinciding with the Paris climate agreement, the launch of the UN Sustainable Development Goals and large asset owners becoming signatories of the UN’s PRI. Along with the surging interest, there come many questions surrounding how best to assess companies from the ESG perspectives. The conventional way is to rely mainly on companies’ disclosed information and index providers such as MSCI to model companies and derive ESG ratings or scores. However, this approach has several drawbacks such as lack of up-to-date assessments on companies because the reporting frequency is typically annual, no common reporting standards globally that mandate companies to disclose information in a consistent way, and human analysis is slow and hard to scale.

The advancements and applications of artificial intelligence and machine learning have disrupted many industries in terms of the ways they operate, and ESG is no exception. The rise of AI in ESG valuation is evident in the number of such vendors coming to the market, offering ingestion of broad sources of data in order to assess companies in a more timely and objective manner.

In our report, we have selected three vendors which utilise AI / machine learning techniques in different ways according to their own philosophies to assess companies’ ESG performance. What they have in common is ingesting large amounts of data and evaluating companies from an outside-in perspective. While the ESG criteria they employ to generate their own scores and rating might be different, there is a clear trend of convergence towards SASB and UN SDGs as the standard way of assessing companies. This could mark the beginning of having a widely accepted common reporting standard, since such convergence appears to be driven by investor demand.

Also, the fact that these vendors utilise AI / machine learning to score companies means that they are able to expand their universe and scale much more efficiently, so long as they have trustworthy and large volumes of data sources they can rely on. This is in stark contrast to the analyst-driven model that traditional ESG score providers adopt. The analyst-driven model would be more prone to subjectivity while the AI-led ESG vendors utilise analyst expertise at the beginning of the process to ensure companies are assessed in a consistent manner.

Another issue with the analyst-driven model is the lack of transparency as to how the scores are derived. Through our discussions with RepRisk, Truvalue and Arabesque S-Ray, they appear to offer greater transparency in how they utilise their data to arrive at the ESG scores and rating. In particular, with RepRisk's and Truvalue's platforms, users can drill down to the specific articles or reports that lead to the final rating or score. The table in Figure 87 summarises the commonalities and differences amongst RepRisk, Truvalue and Arabesque S-Ray.

---

Figure 87. Summary of RepRisk, Truvalue and Arabesque S-Ray Offerings

<table>
<thead>
<tr>
<th>ESG Screening</th>
<th>RepRisk</th>
<th>Positive &amp; Negative</th>
<th>Positive &amp; Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Employees</td>
<td>120 globally with 70 ESG analysts</td>
<td>50 globally with 16 ESG experts</td>
<td>28 globally with 9 in ESG research</td>
</tr>
<tr>
<td>Data Sources</td>
<td>80,000+</td>
<td>100,000+</td>
<td>250+ ESG Metrics sourced from third party vendors who utilise 50,000+ datasets</td>
</tr>
<tr>
<td>Data History</td>
<td>From 2007</td>
<td>From 2007</td>
<td>From 2004</td>
</tr>
<tr>
<td>Languages Covered</td>
<td>20</td>
<td>7 (from July 2019 onwards)</td>
<td>4</td>
</tr>
<tr>
<td>ESG Criteria</td>
<td>Their own 28 ESG issues and 57 ESG topics, currently mapping to SASB materiality map and UN SDGs</td>
<td>SASB Materiality Categories, have been mapped to UN SDGs</td>
<td>Their own 22 sustainability topics, SASB Materiality Map used in their baseline materiality weight calculation, currently developing SDG Quantification Tool</td>
</tr>
<tr>
<td>AI / Machine Learning</td>
<td>NLP, proprietary algorithm to quantify risks</td>
<td>NLP, proprietary algorithm to identify, categorise, quantify and produce ESG scores and analytics</td>
<td>NLP used by their news-based and NGO-based data providers. Semi-supervised dimensionality reduction technique used in-house to structure input data along the 22 topics. Supervised machine learning used to find material features that explains future performance</td>
</tr>
<tr>
<td>Scores</td>
<td>RepRisk Index (RRI), RepRisk Rating (RRR), UNGC Violation Flags</td>
<td>Pulse, Insight, Momentum and Volume</td>
<td>ESG score, E/S/G individual scores, GC Scores, Preferences Filter</td>
</tr>
<tr>
<td>Transparency</td>
<td>Can drill down to event details</td>
<td>Can drill down to detailed data level</td>
<td>Able to drill down the in-house developed score but limited on the data as it is sourced from third party vendors</td>
</tr>
</tbody>
</table>

Source: RepRisk, Truvalue, Arabesque S-Ray, CGDI
Putting the methodology and construction differences aside, we have also shown in our empirical analysis that the ESG scores from these AI vendors could add alpha to investment processes and could complement the scores derived by analyst-driven vendors, especially when external data is limited due to lack of media attention on ESG issues in local markets or languages not supported by AI vendors.

Another consideration is investment horizon as AI-led vendors strive to provide up-to-date ESG performance of companies as opposed to analyst-driven ESG scores and ratings which are for longer-term by design. We have seen asset managers increasingly use the combination of both ESG sources and tilt the weighting of one against the other to suit the investment horizons of their strategies. This approach has the advantage of lower turnover while keeping a close eye on significant ESG issues that could potentially affect the companies they hold.

In summary, from our discussions and analysis conducted on the three AI ESG vendors, we think they present an interesting alternative way to assess ESG performance in addition to analyst-driven ESG providers. There are certainly still issues that remain to be resolved such as coverage in EM but their ease of scalability and adaptability means they can expand universes and adjust to new standards.

This study is set out to provide an overview of AI-led ESG providers and compare them to analyst-driven ESG vendors. There are many features available in the datasets from the three AI-led vendors we have chosen which we have not yet explored in-depth. A deep-dive analysis of such features from each vendor would be in scope as our future projects.
Market Commentary Disclaimer

This communication has been prepared by individual sales and/or trading personnel of Citigroup Inc. or its subsidiaries or affiliates (collectively Citi). In the United Kingdom: Citigroup Global Markets Limited (“CGML”) is authorised by the Prudential Regulation Authority and regulated by the Financial Conduct Authority and the Prudential Regulation Authority (together, the UK Regulator) and has its registered office at Citigroup Centre, Canada Square, London E14 5LB. Amongst its affiliates, (i) Citibank, N.A., London Branch is authorised and regulated by Office of the Comptroller of the Currency (USA), authorised by the Prudential Regulation Authority and subject to regulation by the Financial Conduct Authority and limited regulation by the Prudential Regulation Authority and has its UK establishment office at Citigroup Centre, Canada Square, London E14 5LB and (ii) Citibank Europe plc, UK Branch is authorised by the Central Bank of Ireland and by the Prudential Regulation Authority and subject to regulation by the Central Bank of Ireland, and limited regulation by the Financial Conduct Authority and the Prudential Regulation Authority and has its UK establishment office at Citigroup Centre, Canada Square, London E14 5LB. Outside the UK: i) Citibank Europe plc (“CEP”) is Licensed by the European Central Bank and regulated by the Central Bank of Ireland and the European Central Bank under the Single Supervisory Mechanism and has its registered office at 1 North Wall Quay, Dublin 1, ii) Citibank Europe plc branches located in the EEA are subject to regulation by the respective host country regulator and the Central Bank of Ireland (iii) Citigroup Global Markets Europe AG (“CGME”), authorised and regulated by the Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin) and has its registered office at Reuterweg 16, 60323 Frankfurt am Main, Germany This communication is directed at persons (i) who have been or can be classified by Citi as eligible counterparties or professional clients in line with applicable rules, (ii) Persons in the United Kingdom, who have professional experience in matters relating to investments falling within Article 19(1) of the Financial Services and Markets Act 2000 (Financial Promotion) Order 2005 and (iii) other persons to whom it may otherwise lawfully be communicated. No other person should act on the contents or access the products or transactions discussed in this communication. In particular, this communication is not intended for retail clients and Citi will not make such products or transactions available to retail clients. The information contained herein may relate to matters that are not regulated by any applicable financial services regulatory body, and not subject to protections under any relevant law including protection under any applicable financial services compensation scheme.

To the extent that this communication/these materials has/have been produced in the UK by CGML or Citibank N.A. London branch, it is/they are intended for distribution solely to clients of Citi in jurisdictions where such distribution is permitted and the recipient shall not provide or distribute such materials to any person located in a jurisdiction where it would otherwise trigger a financial services licensing requirement.

To the extent that this communication/these materials has/have been produced by CGME, it is/they are intended for distribution solely to clients of Citi in jurisdictions where such distribution is permitted and the recipient shall not provide or distribute such materials to any person located in a jurisdiction where it would otherwise trigger a financial services licensing requirement.

Citi may be the issuer of, or may trade as principal in, the products or transactions referred to in this communication or other related products or transactions. The sales and/or trading personnel involved in preparing or issuing this communication (the Author) may have discussed the information contained in this communication with others within Citi (together with the Author, the Relevant Persons) and the Relevant Persons may have already acted on the basis of this information (including by trading for Citi’s principal accounts or communicating the information contained herein to other customers of Citi). Certain personnel or business areas of Citi may have access to or have acquired material non-public information that may have an impact (positive or negative) on the information contained herein, but that is not available to or known by the Author of this communication.

Citi is not acting as your agent, fiduciary or investment adviser and is not managing your account. The provision of information in this communication is not based on your individual circumstances and should not be relied upon as an assessment of suitability for you of a particular product or transaction. It does not constitute investment advice and Citi makes no recommendation as to the suitability of any of the products or transactions mentioned. Even if Citi possesses information as to your objectives in relation to any transaction, series of transactions or trading strategy, this will not be deemed sufficient for any assessment of suitability for you of any transaction, series of transactions or trading strategy. Save in those jurisdictions where it is not permissible to make such a statement, we hereby inform you that this communication should not be considered as a solicitation or offer to sell or purchase any securities, deal in any product or enter into any transaction. You should make any trading or investment decisions in reliance on your own analysis and judgment and/or that of your independent advisors and not in reliance on Citi and any decision whether or not to adopt any strategy or engage in any transaction will not be Citi’s responsibility. Citi does not provide investment, accounting, tax, financial or legal advice; such matters as well as the suitability of a potential transaction or product or investment should be discussed with your independent advisors. Prior to dealing in any product or entering into any transaction, you and the senior management in your organisation should determine, without reliance on Citi, (i) the economic risks or merits, as well as the legal, tax and accounting characteristics and consequences of dealing with any product or entering into the transaction (ii) that you are able to assume these risks, (iii) that such product or transaction is appropriate for a person with your experience, investment goals, financial resources or any other relevant
circumstance or consideration. Where you are acting as an adviser or agent, you should evaluate this communication in light of the circumstances applicable to your principal and the scope of your authority.

The information set forth in this communication is not intended to be used for the purposes of (a) determining the price or amounts due in respect of one or more financial products, and/or (b) measuring or comparing the performance of a financial product or a portfolio of financial instruments, and any such use is strictly prohibited without the prior written consent of Citi.

The information in this communication, including any trade or strategy ideas, is provided by individual sales and/or trading personnel of Citi and not by Citi’s research department and therefore the directives on the independence of research, and rules prohibiting dealing ahead of dissemination, do not apply. Any view expressed in this communication may represent the current views and interpretations of the markets, products or events of such individual sales and/or trading personnel and may be different from other sales and/or trading personnel and may also differ from Citi’s published research – the views in this communication may be more short term in nature and liable to change more quickly than the views of Citi research department which are generally more long term. On the occasions where information provided includes extracts or summary material derived from research reports published by Citi’s research department, you are advised to obtain and review the original piece of research to see the research analyst’s full analysis. Any prices used herein, unless otherwise specified, are indicative. Although all information has been obtained from, and is based upon sources believed to be reliable, it may be incomplete or condensed and its accuracy cannot be guaranteed. Citi makes no representation or warranty, expressed or implied, as to the accuracy of the information, the reasonableness of any assumptions used in calculating any illustrative performance information or the accuracy (mathematical or otherwise) or validity of such information. Any opinions attributed to Citi constitute Citi’s judgment as of the date of the relevant material and are subject to change without notice. Provision of information may cease at any time without reason or notice being given. Commissions and other costs relating to any dealing in any products or entering into any transactions referred to in this communication may not have been taken into consideration.

Any scenario analysis or information generated from a model is for illustrative purposes only. Where the communication contains “forward-looking” information, such information may include, but is not limited to, projections, forecasts or estimates of cash flows, yields or return, scenario analyses and proposed or expected portfolio composition. Any forward-looking information is based upon certain assumptions about future events or conditions and is intended only to illustrate hypothetical results under those assumptions (not all of which are specified herein or can be ascertained at this time). It does not represent actual termination or unwind prices that may be available to you or the actual performance of any products and neither does it present all possible outcomes or describe all factors that may affect the value of any applicable investment, product or investment. Actual events or conditions are unlikely to be consistent with, and may differ significantly from, those assumed. Illustrative performance results may be based on mathematical models that calculate those results by using inputs that are based on assumptions about a variety of future conditions and events and not all relevant events or conditions may have been considered in developing such assumptions. Accordingly, actual results may vary and the variations may be substantial. The products or transactions identified in any of the illustrative calculations presented herein may therefore not perform as described and actual performance may differ, and may differ substantially, from those illustrated in this communication. When evaluating any forward looking information you should understand the assumptions used and, together with your independent advisors, consider whether they are appropriate for your purposes. You should also note that the models used in any analysis may be proprietary, making the results difficult or impossible for any third party to reproduce. This communication is not intended to predict any future events. Past performance is not indicative of future performance.

Citi shall have no liability to the user or to third parties, for the quality, accuracy, timeliness, continued availability or completeness of any data or calculations contained and/or referred to in this communication nor for any special, direct, indirect, incidental or consequential loss or damage which may be sustained because of the use of the information contained and/or referred to in this communication or otherwise arising in connection with the information contained and/or referred to in this communication, provided that this exclusion of liability shall not exclude or limit any liability under any law or regulation applicable to Citi that may not be excluded or restricted.

Citi may submit prices, rates, estimates or values to data sources that publish indices or benchmarks which may be referenced in products or transactions discussed in this communication. Such submissions may have an impact on the level of the relevant index or benchmark and consequently on the value of the products or transactions. Citi will make such submissions without regard to your interests under a particular product or transaction. Citi has adopted policies and procedures designed to mitigate potential conflicts of interest arising from such submissions and our other business activities. In light of the different roles performed by Citi you should be aware of such potential conflicts of interest.

The transactions and any products described herein may be subject to fluctuations of their mark Document to-market price or value and such fluctuations may, depending on the type of product or security and the financial environment, be substantial. Where a product or transaction provides for payments linked to or derived from prices or yields of, without limitation, one or more securities, other instruments, indices, rates, assets or foreign currencies, such provisions may result in negative fluctuations in the value of and
amounts payable with respect to such product prior to or at redemption. You should consider the implications of such fluctuations with your independent advisers. The products or transactions referred to in this communication may be subject to the risk of loss of some or all of your investment, for instance (and the examples set out below are not exhaustive), as a result of fluctuations in price or value of the product or transaction or a lack of liquidity in the market or the risk that your counterparty or any guarantor fails to perform its obligations or, if the product or transaction is linked to the credit of one or more entities, any change to the creditworthiness of the credit of any of those entities.

Citi (whether through the individual sales and/trading personnel involved in the preparation or issuance of this communication or otherwise) may from time to time have long or short principal positions and/or actively trade, for its own account and those of its customers, in products identical to or economically related to the products or transactions referred to in this communication. Citi may also undertake hedging transactions related to the initiation or termination of a product or transaction, that may adversely affect the market price, rate, index or other market factor(s) underlying the product or transaction and consequently its value. Citi may have an investment banking or other commercial relationship with and access to information from the issuer(s) of securities, products, or other interests underlying a product or transaction. Citi may also have potential conflicts of interest due to the present or future relationships between Citi and any asset underlying the product or transaction, any collateral manager, any reference obligations or any reference entity.

Any decision to purchase any product or enter into any transaction referred to in this communication should be based upon the information contained in any associated offering document if one is available (including any risk factors or investment considerations mentioned therein) and/or the terms of any agreement. Any securities which are the subject of this communication have not been and will not be registered under the United States Securities Act of 1933 as amended (the Securities Act) or any United States securities law, and may not be offered or sold within the United States or to, or for the account or benefit of, any US person, except pursuant to an exemption from, or in a product or transaction, not subject to, the registration requirements of the Securities Act. This communication is not intended for distribution to, or to be used by, any person or entity in any jurisdiction or country which distribution or use would be contrary to law or regulation.

Citi may offer, issue, distribute or provide other services (including, without limitation, custodial and other post-trade services) in relation to certain financial instruments. Some of these financial instruments may be unsecured financial instruments issued or entered into by BRRD Entities (i.e. EEA entities within the scope of Directive 2014/59/EU (the BRRD including as such Directive has been on-shored into UK legislation), including EEA credit institutions, certain EEA investment firms and/or their EEA subsidiaries or parents) (BRRD Financial Instruments).

In various jurisdictions (including, without limitation, the UK, the EEA and the United States) national authorities have certain powers to manage and resolve banks, broker dealers and other financial institutions (including, but not limited to, Citi) when they are failing or likely to fail. There is a risk that the use, or anticipated use, of such powers, or the manner in which they are exercised, may materially adversely affect (i) your rights under certain types of unsecured financial instruments (including, without limitation, BRRD Financial Instruments), (ii) the value, volatility or liquidity of certain unsecured financial instruments (including, without limitation, BRRD Financial Instruments) that you hold and / or (iii) the ability of an institution (including, without limitation, a BRRD Entity) to satisfy any liabilities or obligations it has to you. You may have a right to compensation if the exercise of such powers results in less favourable treatment for you than the treatment that you would have received under normal insolvency proceedings. By accepting any services from Citi, you confirm that you are aware of these risks. Some of these risks (in particular the risks that arise under the BRRD) are set out in more detail at the link below and you are deemed to have reviewed and considered such risks prior to any decision to purchase any product or enter into any transaction referred to in this communication.

This communication contains data compilations, writings and information that are confidential and proprietary to Citi and protected under copyright and other intellectual property laws, and may not be reproduced, distributed or otherwise transmitted by you to any other person for any purpose unless Citi’s prior written consent have been obtained.


In any instance where distribution of this communication is subject to the rules of the US Commodity Futures Trading Commission (“CFTC”), this communication constitutes an invitation to consider entering into a derivatives transaction under U.S. CFTC Regulations §§ 1.71 and 23.605, where applicable, but is not a binding offer to buy/sell any financial instrument.

**IRS Circular 230 Disclosure:** Citigroup Inc. and its affiliates do not provide tax or legal advice. Any discussion of tax matters in these materials (i) is not intended or written to be used, and cannot be used or relied upon, by you for the purpose of avoiding any tax penalties and (ii) may have been written in connection with the “promotion or marketing” of a transaction (if
relevant) contemplated in these materials. Accordingly, you should seek advice based your particular circumstances from an independent tax advisor.

Although CGML and CGME are affiliated with Citibank, N.A. (together with Citibank, N.A.’s subsidiaries and branches worldwide, Citibank), you should be aware that none of the products mentioned in this communication (unless expressly stated otherwise) are (i) insured by the Federal Deposit Insurance Corporation or any other governmental authority, or (ii) deposits or other obligations of, or guaranteed by, Citibank or any other insured depository institution.

CGML’s operations are not subject to the supervision of the Israel Securities Authority. The permit granted by the Israeli Securities Authority does not constitute an opinion regarding the quality of the services rendered by the permit holder or the risks that such services entail.

© 2019 Citigroup Global Markets Limited. Citi, Citi and Arc Design are trademarks and servicemarks of Citigroup Inc. or its affiliates and are used and registered throughout the world.