The Ubiquity of Data: **Challenges and Opportunities for Asset Managers**

A swell in the number of digital information sources, coupled with the rapid growth in storage capacity, computational power, and connectivity, has led to data being created and processed on unprecedented scales. Seven decades into this global digital revolution, investors have a wealth of data to parse and manipulate with the aim of extracting pertinent information for broad investment application—from security selection through to portfolio design and customisation. Asset managers with specialist capabilities in the analysis and processing of big data sets are better placed to extract actionable information from otherwise nebulous sources, unlocking their potential to generate returns, enhance capital allocation decisions, and optimise portfolios.

LAZARD
ASSET MANAGEMENT

The sprawling nature of the internet, the growing number of interconnected devices, and the rapid pace of automation mean that data creation will continue to rise at a colossal rate. Not only is there an expansion in the volume of data being generated, but the velocity with which it is transmitted and the variety of forms that data may take are also multiplying. Structured and unstructured data sets can take a number of different forms, encompassing traditional data, alternative data, and big data (Exhibit 1).

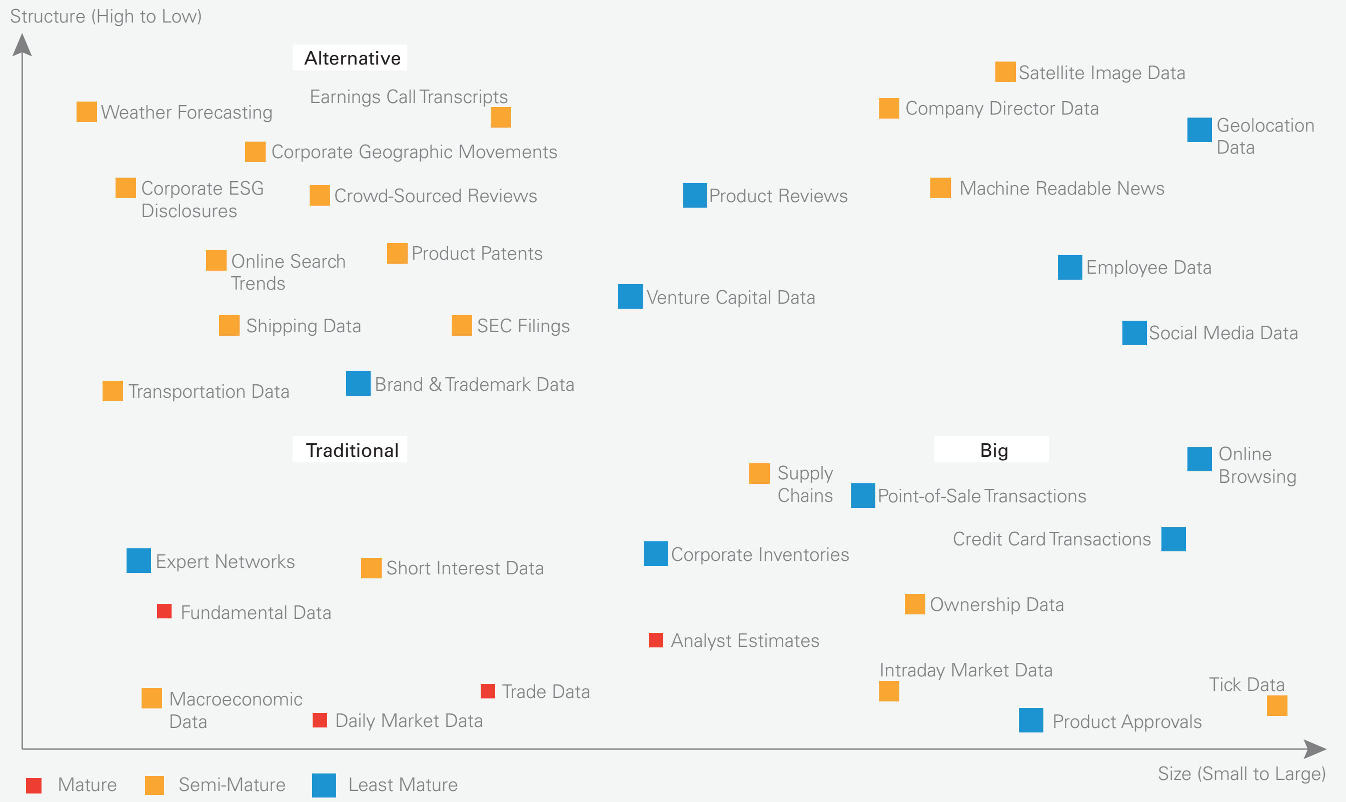
The field of data science is concerned with extracting useful insights from data, irrespective of the form the data takes. Although data science was first coined as a term in the mid-2000s, the discipline is essentially an extension and convergence of existing methods. It draws on advancements in the fields of statistics, computer science, data technology, visualisation methods, and mathematics that have origins tracing back well over a century. As such, we believe that the analysis of big data marks a natural evolutionary step forward in data analysis rather than a revolutionary shift (Exhibit 2), although in some areas, such as artificial intelligence (AI) and machine learning (ML), the change in scale in computing and data has been revolutionary in terms of real-world consequences.

In contrast to traditional methods, which slot data into pre-specified relationships, AI is able to act on data much as the human mind does, by reacting to data and interpreting it in order to test a hypothesis. AI is built on an iterative ML process that results in the analysis of data being continually refined and improved upon, such that the ML component adapts its behaviour based on changes in findings so that any interpretations drawn from similar data in the future contain fewer errors. Big data is fundamental to this process as it feeds the algorithms that drive the ML. Relationships are automatically discovered and adapted as more data is ingested, so the quality and quantity of data are both critically important.

Handle with (Specialist) Care

While data processing techniques continue to grow in sophistication, requiring increasingly specialised handling, the significance of the relationships inferred by them heavily rests on the integrity of the data supplied and the rationale behind its selection. That requires investors to set out clearly defined objectives around what they hope to achieve by using the data—including the investment hypotheses and causal mechanisms that they seek to test—followed by rigorous data validation. Ongoing

Exhibit 1
New Forms of Data Tend to Be Less Readily Digestible and Voluminous

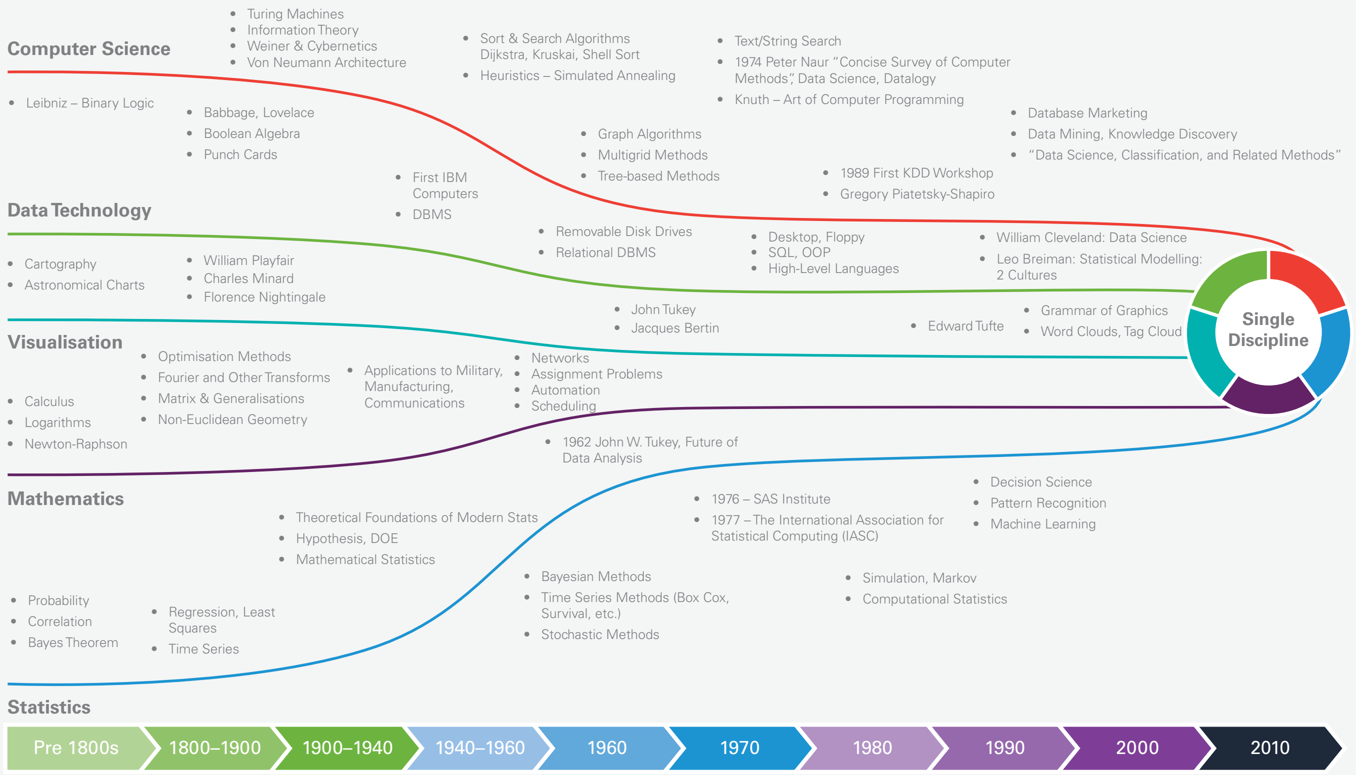


As at 30 April 2019

For illustrative purposes only. Data sets represented by red points are examples of the most mature items available, typically first accessible over 40 years ago and ubiquitous in their usage. Data sets represented in orange are less mature, with more recent availability (typically first gathered between 10 and 20 years ago) and requiring a degree of specialised processing. Data sets represented in blue are the least mature in availability, representing the latest developments in data capture and often requiring novel techniques for effective processing.

Source: Lazard

Exhibit 2
Modern Data Science Techniques Have Roots in Traditional Sciences



For illustrative purposes only
Source: Capgemini

data validation and careful data handling provide the bedrock of the ML process. Flaws in either can severely undermine the significance of the relationships determined by the algorithms within the AI framework, and potentially lead investors to form misguided interpretations from them. Data integrity validation should span multiple considerations, including:

- Data vendor risk—including risks around credibility, reputation, and continuity (i.e., the risk of the vendor ceasing to exist or provide a set of data)
- Data provenance—including ethical implications around how data is sourced, stored, shared, and used
- Material non-public information risk—the act of accessing data in itself does not necessarily constitute it as public information
- Data integrity—sparsity, completeness, accuracy, and bias assessments
- Data chain and ownership—including considerations around intellectual property and ownership rights of the insights derived from data

Given these considerations, data validation and preparation is a necessarily holistic process, requiring diligent screening and handling to ensure that biases within data are correctly accounted for, inaccuracies addressed and, if possible, gaps in information appropriately filled.

Beyond the detailed preparation of data, rigorous model training[1] and diligent monitoring are the next critical components of the ML process—ensuring that the data analysis techniques are well suited to the task at hand, as well as being mathematically and computationally valid. However, this still may not prevent spurious correlations and false inferences. Data validation and model training might make the ML architecture and infrastructure sound, but the underlying causal links between the inputs fed through them and the outputs generated by them could remain opaque, rendering the system a "black box." Discipline-specific knowledge, or domain knowledge, is imperative in ensuring that the relationships found are causal and intelligible. Data scientists often have a strong background in the mathematical sciences but may have limited expertise in economic and financial theory or access to specialists. As such, the relationships discovered by the ML models may lack context, leading to sub-optimal or even invalid conclusions.

Domain Knowledge, the Light in Data Analytics

Collaboration between investment professionals and data science specialists can serve to complement investment decision-making processes—first, by formulating detailed hypotheses that can be tested and by identifying the most relevant and valuable data sets to apply, and second, by building confidence in the model output. We believe that our data scientists are well positioned in this regard as they work alongside our fundamental research analysts, giving them access to a proven investment framework within which to validate the relationships generated by ML models. This ensures there is conviction behind investment ideas and that any conclusions drawn from data are based on statistical rigour and rooted in a sound investment rationale. Additionally, our data scientists' expertise originates from broad academic backgrounds, including astrophysics, economics, computer science, and medical imaging. In our view, these diverse skill sets bolster the data validation and investment research processes by offering differentiated perspectives to complex investment problems beyond that offered by the traditional realm of mathematical sciences.

Very few investment management companies have been able to successfully apply and scale the innovation and technology in AI to generate positive firm-wide impact, despite overly ambitious promises. As such, the true potential of data science is often not met with real-life practical implementation. We believe data science should be approached with a healthy dose of scepticism rather than being viewed as a panacea that is able to remedy any strategic or investment challenge. We believe that this philosophy has meant that we are able to truly demonstrate with confidence the areas in which data science has value and where it is able to work best in a practical sense. As the field of data science has grown, we have carefully considered how to harness the opportunities alternative data offers more broadly across the firm, fostering deeper collaboration between teams and—through the use of proprietary technology—internally socialising research efforts, and sharing and integrating insights to enhance investment decision-making processes. By carefully combining techniques at the forefront of data science with detailed fundamental stock-level assessments, we believe the information edge derived from alternative data sources can be translated into enhanced investment insights.

What's the Big Idea? Finding Practical Value from Alternative Data

For almost a decade, Lazard's Multi-Asset and Quantitative Equity teams have looked at how alternative data can complement their investment and decision-making processes. Below, we highlight recent research projects undertaken by the teams in close collaboration with the firm's data scientists.

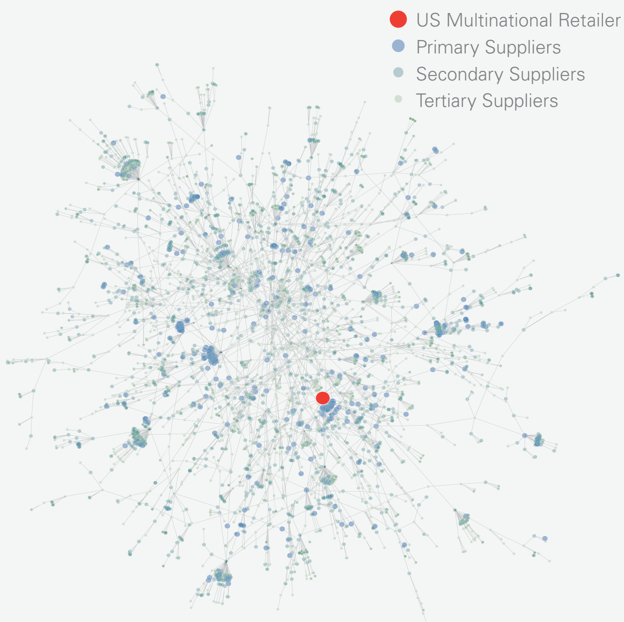
Network Analysis to Model ESG Risk Propagation

ESG (Environmental, Social, and Governance) ratings are useful tools that aid our investment analysis and security selection decisions. Third-party ratings companies tend to assess whether a company's operational and financial activities undermine or strengthen its ability to manage industry-specific ESG risks relative to peers when generating an ESG score. However, the ESG risks faced by any given company are likely to stretch beyond its designated industry. Highly complex supply chains—which often span multiple countries—create an interdependent network of companies in which ESG risk factors are closely intertwined across sectors and through different levels of the supply chain. Supply chain considerations tend not to be reflected by the ESG scores of third-party providers, highlighting the importance for asset managers to view ESG issues holistically from multiple perspectives by conducting their own research.

We are able to demonstrate how information from a variety of sources, including corporate filings, companies' press releases, and investor presentations, can be handled using data science techniques to map complex supply chain relationships. The transfer of ESG risk can be modelled by propagating E, S, and G scores through the network in an iterative fashion, depending on the strength of the relationship between various companies and their respective suppliers.

By propagating ESG risks through this network model, we find that a company's ESG scores can be materially affected by its suppliers and distributors, resulting in assessments that deviate from the original score. As such, we believe there is merit in considering a company's supply chain when determining how sustainable that company is from a holistic ESG perspective. As an example, when applying this model to a US multinational retailer, the scale of the company's extensive reach, and the companies within its comprehensive supply chain network, becomes apparent (Exhibit 3).

Exhibit 3
ESG Risks Are Able to Propagate through Supply Chain Networks



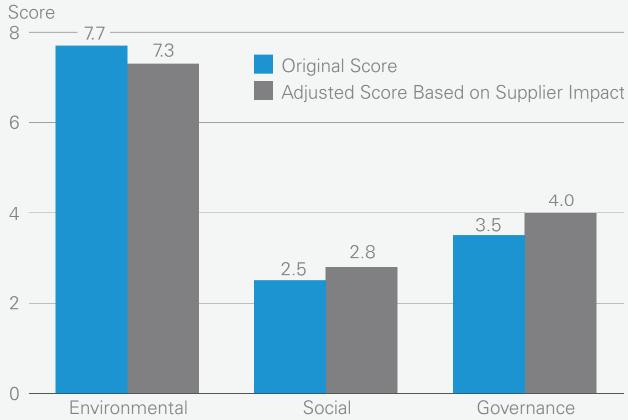
As at 31 December 2018

For illustrative purposes only

A minimum spanning tree representation of the supply chain network for suppliers connected to a widely known US multinational retailer, truncated to companies within three degrees of separation. This plot contains 4,808 companies (including the central company under observation), but could be extended to all degrees of separation. The network model modifies standalone ESG scores to propagate ESG risk through the network using quantitative methods, self-consistently taking account of the complex interconnected relationships among companies within the network.

Source: Lazard, FactSet, MSCI

Exhibit 4
Supply Chain Analysis Can Offer Different Assessments of ESG Risks at the Company Level



As at 31 December 2018

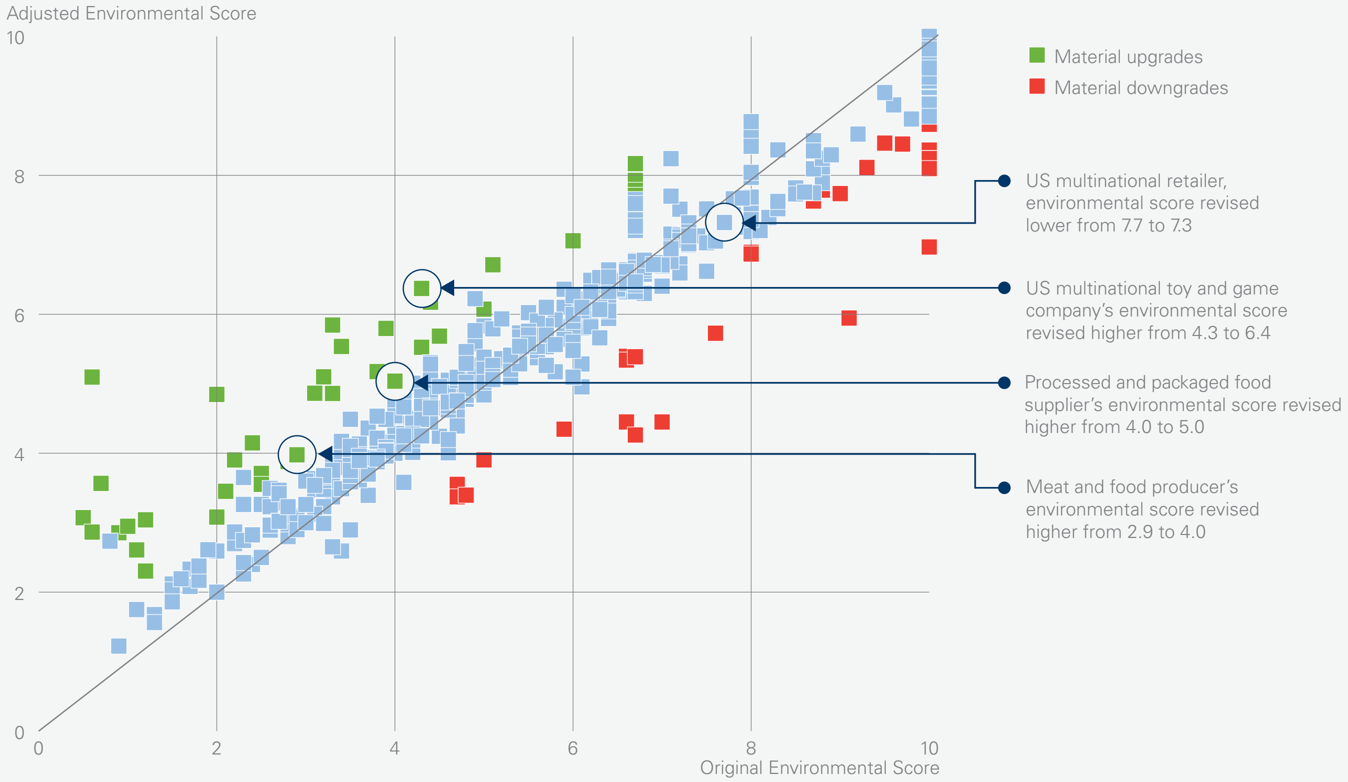
Modification of standalone environmental, social, and governance scores for the US multinational retailer, accounting for the propagated impact of ESG credentials of its supplier network. Scores range from 0 to 10, with 0 being the weakest ESG score and 10 the strongest.

Source: Lazard, FactSet, MSCI

We adopt several techniques from network theory when modelling the propagation of ESG scores within this complex network. When factoring in the influence of direct or indirectly linked suppliers and distributors on the central subject—the US retailer—our analysis leads to an adjustment of the retailer's original ESG scores (Exhibit 4). Although the adjustments are modest for this company (we could consider a material change to be greater than or equal to ±1), monitoring scores to track changes may prove to be prudent should any of them shift significantly, owing to supplier network effects.

Indeed, when looking at adjustments to environmental scores across all companies within the network in the same way, more than 1 in 10 companies experience a material adjustment to their environmental score, either to the upside or downside, when supplier impact is taken into account (Exhibit 5). As the supply chain network is an ecosystem of direct and indirectly linked companies, material adjustments to ESG scores could feed back to the central company under observation over time, in this case, the US multinational retailer.

Exhibit 5
Supply Chain Analysis Can Offer Different Assessments of ESG Risks across Entire Networks



As at 31 December 2018

The impact of environmental risk propagation for all companies within the US multinational retailer's supplier network, based on their supply chain network properties. The plot shows adjusted environmental scores for each company, based on impact from suppliers within their network, against each company's original environmental score (i.e., those derived from third-party sources that do not take supply chain analysis into account). Scores may be revised higher or lower relative to their original levels. Scores range from 0–10, with 0 being the weakest environmental score and 10 the strongest. Companies with a material upward adjustment to their environmental score (defined as a change that is greater than or equal to +1) shown in green. Companies with a material downward adjustment to their environmental score (defined as a change that is greater than or equal to -1) shown in red. 14% of the companies in the network experienced a material adjustment to their environmental score (8% of companies were adjusted upward, while 6% of companies were adjusted downward). The three suppliers specifically mentioned have a relationship value (defined as a percentage revenue dependency from a supplier multiplied by the supplier's LTM revenue) greater than $500 million in relation to the central company under observation.

Source: Lazard, FactSet, MSCI

While we believe there is merit to using third-party ESG data, as a firm we think there is a growing need to critically evaluate industry findings with in-house insights—through quantitative methods, fundamental analysis, or a combination of both—and we continue to look for ways to further enhance and refine our ESG assessments through such approaches.

Natural Language Processing to Derive Signals from Earnings Call Transcripts

Text-based sentiment measures have the potential to provide unique insights into companies which would otherwise be difficult to determine. The word choice, complexity of language, and tone used during an earnings call can offer insights into a broad range of financially material factors that cannot be gleaned from financial reports—how a company's culture is evolving, for instance, and insights into the personality traits of the senior figures at the firm which could offer clues about the company's direction. By tracking the evolution of natural language processing (NLP) scores, we can develop a deeper understanding of a company's trajectory in absolute terms and relative to industry peers.

We have produced a multi-faceted trading factor based on several hypotheses relating company performance to the tone, complexity, and topics discussed during earnings call events by isolating a number of features that we believe to be useful in generating quantitative trading signals from this rich data source. By way of example, we highlight an earnings call for a US automotive company, highlighting the differences in processed information extracted for different call participants within different parts of the call (Exhibit 6).

Big Data's Role in Solutions Customisation

AI not only offers scope to enhance security analysis and selection and improve investment processes, but it is also helping with the sophisticated customisation of portfolios through highly targeted asset allocation decisions. Investors are becoming more focused on building portfolios that embody their investment objectives and values. In this process, they have traditionally been constrained by asset class or geography as the starting point for portfolio construction. Now, through alternative data and ML, they are starting to deconstruct and re-categorise asset choices. These techniques enable them to transcend conventional categories, build bespoke portfolios, and free themselves from the characteristic mapping that occurs around traditional asset allocation decisions.

Lazard's Multi-Asset and Quantitative Equity teams along with the firm's in-house data scientists have created a proprietary platform that incorporates quantitative and fundamental factors to design portfolios from the bottom up, free of traditional constraints, starting with the client's objective. The platform is able to conform stocks to any desired configuration—based on what the client seeks to achieve—to create portfolios that exploit market segments in a way traditional portfolio construction likely could not. This might include allocating capital to the area of robotics and automation—companies in the investable universe that offer products and services throughout the robotics and automation value chain—or even companies in the "internet of things" ecosystem. Currently, no single sector or industry exists that spans these opportunities, highlighting the need to break down categorisation concepts in order to create truly bespoke investment solutions.

Exhibit 6

Earnings Call Transcripts Offer a Window to Otherwise "Hidden" Data

	Sections		People		
	Prepared Statements	Unscripted Q&A	Company Management	External Analysts	All Sections
Number of Sentences	18	189	55	152	219
Number of Words	1,098	10,305	7,804	3,599	11,403
Word Sentiment	0.67	-0.02	0.69	-0.18	0.28
Sentence Sentiment	0.64	-0.06	0.72	-0.23	0.32
Topic Sentiment	0.00	-0.13	0.17	-0.38	-0.11
Language Complexity	8.9	6.3	8.1	5.8	7.8

The earnings call narrative was dominated by company insider dialogue during the Q&A portion, suggesting long-winded answers from management may have been used to avoid answering questions directly

Sentiment during the Q&A section was mildly negative when compared to the prepared statements, suggesting a degree of linguistic vetting by company management

Management uses marginally more complex language than analysts, which could suggest obfuscation

Alternative definitions of sentiment may yield contrasting views on a particular piece of text, helping to build a more comprehensive assessment of the language being used, generating richer analysis

As at 31 December 2018

An example of the quantitative information extracted from a particular earnings call. By dividing the call into various components (people and sections) we achieve a higher level of granularity within the analysis. We are also able to extract a variety of attributes, from basic quantities relating to the length of the document to the complexity, sentiment, and grammatical structure of the narrative. This structured output consequently lends itself to further quantitative processing, such as factor construction.

Source: FactSet

Sector and sub-industry categorisations are one limiting factor, the size of the investable universe they create can be another. If investors wanted a custom portfolio designed to gain exposure to, for instance, timber, traditional classifications would confine them to the materials sector—namely the containers and packaging, and paper and forest products sub-industries. Such companies would be involved with the production of paper, paper-based containers and packaging, and the production of timber, lumber, and other wood products. However, within these categories, there may be too few companies directly linked to timber from which to construct a portfolio. Specialist fundamental investors with intimate knowledge of the timber value chain would be able to intuit related companies in other sectors, such as home-improvement retailers. But when doing this at a global level it is possible to overlook opportunities—the MSCI ACWI Investable Market Index contains 8,675 constituents.[2]

Alternative data sets offer a different starting point from which to construct a comprehensive universe. Financial statements, earnings call transcripts, news data, industry reports, and the analyst notes of "pure" timber stocks can be parsed by using text mining techniques to understand which words and terms correlate strongly with timber companies (Exhibit 7). The words and terms that feature often are ranked more highly and then cross-referenced against stocks from other sectors to map connected companies within the timber value chain (Exhibit 8).

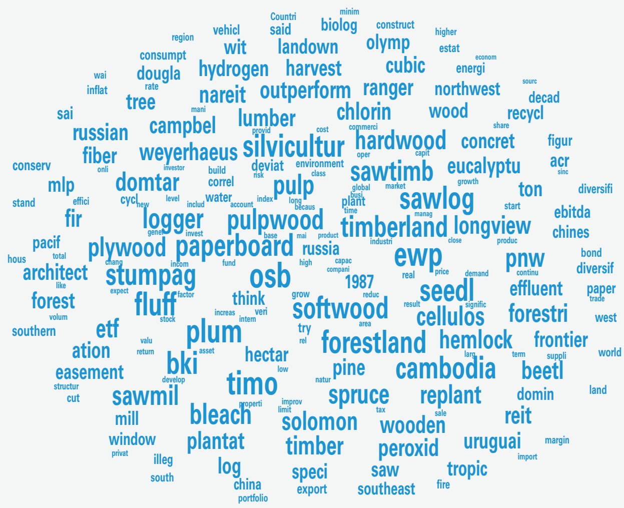
Once the value chain has been defined, the universe is optimised by applying fundamental analysis and quantitative techniques to validate the ML mapping—identifying the causal links that connect the underlying companies and eliminating spurious correlations and false inferences. The fundamental framework and quantitative inputs also allow for the integration of ESG factors through the stock selection process, depending on the investor's requirements. In the case of timber, this could include considerations around timberland management from an environmental and societal perspective, the impact on or of climate change, and whether activities are conservation-based, (e.g., by improving carbon sequestration rates, and through habitat restoration and maintenance). Mapping a universe of stocks in this way, irrespective of country of domicile, categorisation, or market capitalisation, broadens the opportunity set and allows for it to be filtered systematically to curate the desired portfolio.

Adding a New Dimension to a Traditional Investment Edge

An information edge is the cornerstone of successful alpha generation. Whether an information edge is garnered through on-the-ground assessments of company fundamentals, supply chain dynamics, or structural themes—or, indeed through the systemisation of insights gathered from diverse alternative data—these inputs collectively serve to build a fuller picture of a company's long-term prospects than any of these factors would if viewed in isolation. Therefore, alternative data serves as a novel input for investors to use alongside more traditional information sources, and its considerable expansion is creating a new dimension for information asymmetry to occur.

At Lazard, we recognise the alpha-generating opportunity created by the growth in alternative data and see value in its broader application through the portfolio construction and optimisation process. As a firm, we are engaged with understanding how to source, manipulate, and leverage these inputs to reach better investment decisions more quickly and consistently—and ultimately to meet our clients' objectives.

Exhibit 7
The DNA of Sector and Industry Classifications

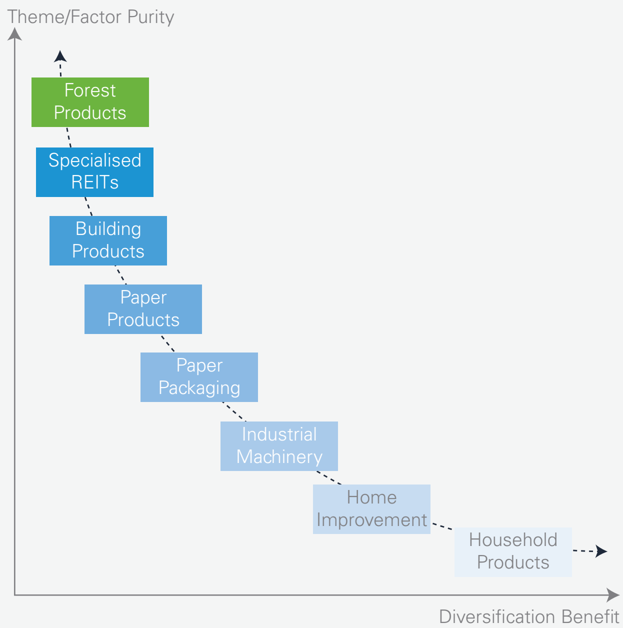


For illustrative purposes only

Word cloud generated using data from SEC filings of companies within related GICS sectors of the MSCI ACWI Investable Market Index. Truncated words are due to "stemming". Stemming consolidates all forms of a single word into one term.

Source: Lazard, MSCI

Exhibit 8
Deconstructing the Timber Value Chain



For illustrative purposes only

Source: Lazard

Acknowledgements

We would like to thank the following people for their contribution to the development of this paper.

Michael Cook MSci, PhD

Vice President, Quantitative Research Analyst
Lazard Asset Management Limited (London)

Michael Cook is a Quantitative Research Analyst in Lazard's Equity Advantage team. He began working in the investment field in 2010. Prior to joining Lazard in 2017, Michael was a Quantitative Analyst at AHL, Man Group, developing equity market neutral strategies across global portfolios. He also had a role within the Product Control team at Lloyds Banking Group, valuing exotic cross-asset derivatives. He has a PhD in Astrophysics from SISSA in Italy and an MSci in Physics with Astronomy (first class honours) from the University of Nottingham.

Jose Perez Sanchez, PhD

Vice President, Data Scientist
Lazard Asset Management LLC (New York)

José Pérez Sánchez is a Data Scientist on the Multi-Asset team. He began working in the investment field in 2016. Prior to joining Lazard in 2017, Jose was a Software Engineer with Bank of America Merrill Lynch. During his career both in academia and in the private sector, he has applied statistical learning techniques to extract information from medical images and performed statistical analyses in the biomedical field. He also worked extensively on numerical simulations and medical image reconstruction and processing. He has a PhD in Physics from Complutense University of Madrid and a Bachelor's Degree in Physics from University of Havana.

Amit Schechter

Vice President, Data Scientist
Lazard Asset Management LLC (New York)

Amit Schechter is a Data Scientist specializing in data visualization and user interface design on the Lazard Multi-Asset investment team. He began working in the investment field in 2018 upon joining Lazard. Prior to joining, Amit Co-founded and headed development at TWO-N, a data visualization agency. Previously he created market data applications and displays at NYSE/SIAC. Amit has taught graphic design and programming at Parsons School of Design and holds a BSAS in Architecture from the University of Illinois at Urbana-Champaign.

Michael Bernadiner

Director, Portfolio Manager/Analyst
Lazard Asset Management LLC (New York)

Michael Bernadiner is a Portfolio Manager/Analyst on the Commodities team. He began working in the investment field in 1998. Prior to joining Lazard in 2016, Michael was a Portfolio Manager for Deutsche Asset Management's Commodity institutional and retail strategies. He has a BS in Finance and Economics from New York University's Leonard N. Stern School of Business.

This content represents the views of the author(s), and its conclusions may vary from those held elsewhere within Lazard Asset Management. Lazard is committed to giving our investment professionals the autonomy to develop their own investment views, which are informed by a robust exchange of ideas throughout the firm.

Notes

1 Model training is a process by which an ML algorithm learns particular relationships in underlying data by being exposed to some representative training data.

2 Source MSCI. As at 29 March 2019

Important Information

Published on 13 May 2019.

Information and opinions presented have been obtained or derived from sources believed by Lazard to be reliable. Lazard makes no representation as to their accuracy or completeness. All opinions expressed herein are as of the published date and are subject to change.

Certain information included herein is derived by Lazard in part from an MSCI index or indices (the "Index Data"). However, MSCI has not reviewed this product or report, and does not endorse or express any opinion regarding this product or report or any analysis or other information contained herein or the author or source of any such information or analysis. MSCI makes no express or implied warranties or representations and shall have no liability whatsoever with respect to any Index Data or data derived therefrom.

This document is provided by Lazard Asset Management LLC or its affiliates ("Lazard") for informational purposes only. Nothing herein constitutes investment advice or a recommendation relating to any security, commodity, derivative, investment management service, or investment product. Investments in securities, derivatives, and commodities involve risk, will fluctuate in price, and may result in losses. Certain assets held in Lazard's investment portfolios, in particular alternative investment portfolios, can involve high degrees of risk and volatility when compared to other assets. Similarly, certain assets held in Lazard's investment portfolios may trade in less liquid or efficient markets, which can affect investment performance. Past performance does not guarantee future results.

This document is intended only for persons residing in jurisdictions where its distribution or availability is consistent with local laws and Lazard's local regulatory authorizations. The Lazard entities that have issued this document are listed below, along with important limitations on their authorized activities.

Australia: Issued by Lazard Asset Management Pacific Co., ABN 13 064 523 619, AFS License 238432, Level 39 Gateway, 1 Macquarie Place, Sydney NSW 2000, which is licensed by the Australian Securities and Investments Commission to carry on a financial services business. This document is intended for wholesale investors only. **Canada**: Issued by Lazard Asset Management (Canada) Inc., 30 Rockefeller Plaza, New York, NY 10112 and 130 King Street West, Suite 1800, Toronto, Ontario M5X 1E3, a registered portfolio manager providing services to non-individual permitted clients. **Dubai**: Issued and approved by Lazard Gulf Limited, Gate Village 1, Level 2, Dubai International Financial Centre, PO Box 506644, Dubai, United Arab Emirates. Registered in Dubai. International Financial Centre 0467. Authorised and regulated by the Dubai Financial Services Authority to deal with Professional Clients only. **EU Member States**: Issued by Lazard Asset Management (Deutschland) GmbH, Neue Mainzer Strasse 75, D-60311 Frankfurt am Main. **Hong Kong**: Issued by Lazard Asset Management (Hong Kong) Limited (AQZ743), One Harbour View Street, Central, Hong Kong. Lazard Asset Management (Hong Kong) Limited is a corporation licensed by the Hong Kong Securities and Futures Commission to conduct Type 1 (dealing in securities) and Type 4 (advising on securities) regulated activities only on behalf of "professional investors" as defined under the Hong Kong Securities and Futures Ordinance (Cap. 571 of the Laws of Hong Kong) and its subsidiary legislation. **Korea**: Issued by Lazard Korea Asset Management Co. Ltd., 10F Seoul Finance Center, 136 Sejong-daero, Jung-gu, Seoul, 100-768. **People's Republic of China**: Issued by Lazard Asset Management. Lazard Asset Management does not carry out business in the P.R.C. and is not a licensed investment adviser with the China Securities Regulatory Commission or the China Banking Regulatory Commission. This document is for reference only and for intended recipients only. The information in this document does not constitute any specific investment advice on China capital markets or an offer of securities or investment, tax, legal or other advice or recommendation or, an offer to sell or an invitation to apply for any product or service of Lazard Asset Management. **Singapore**: Issued by Lazard Asset Management (Singapore) Pte. Ltd., 1 Raffles Place, #15-02 One Raffles Place Tower 1, Singapore 048616. Company Registration Number 201135005W, which provides services only to "institutional investors" or "accredited investors" as defined under the Securities and Futures Act, Chapter 289 of Singapore. **Switzerland**: Issued by Lazard Asset Management Schweiz AG, Usteristrasse 9, CH-8001 Zurich. **United Kingdom**: Issued or approved by Lazard Asset Management Ltd., 50 Stratton Street, London W1J 8LL. Registered in England Number 525667. Authorised and regulated by the Financial Conduct Authority (FCA), providing services only to persons classified as eligible counterparties or professional clients under the rules of the FCA. **United States**: Issued by Lazard Asset Management LLC, 30 Rockefeller Plaza, New York, NY 10112.