17 96

LOMBARD ODIER
INVESTMENT MANAGERS

Big Data is a big deal

An investor's guide to the applications and challenges of alternative data

03/19

March 2019

At a glance

- **Game changer:** Big Data is a relatively new and exciting development that has the potential to re-shape the investing landscape. In this report, we discuss the factors driving growth in Big Data, examine investment use cases and outline key criteria for success.
- **A "Third Way":** For investors used to thinking about investment strategies in a binary - discretionary or systematic – mindset, Big Data offers a fresh alternative that blends best practices from both fields to generate uncorrelated investment insights.
- **New data, new challenges:** Unlike market-centric data sources traditionally used in systematic strategies, Big Data offers insights into real world activity. Translating Big Data into actionable insights requires unique skillsets that represent a significant barrier to entry.
- **Just getting started:** While the ecosystem has grown significantly in recent years, we believe it remains in its nascent stages and expect it to grow meaningfully more complex and voluminous over the near-to medium-term.
- **Keys to success:** Distinct challenges and limitations of Big Data require a tailored investment process. This process must adapt as fast as data evolves to stay ahead. Efficiency with respect to evaluating, refining, and contextualizing data is more paramount than access to scale and resources.

Challenges to successfully using Big Data

p.13

Big Data is changing how we look at investing.

For professional investor use only. Please see important information at the end of this document.

This document has been prepared by:

Gene Getman, CAIA
Client Portfolio Manager and Product Specialist
Lombard Odier Investment Managers (LOIM)

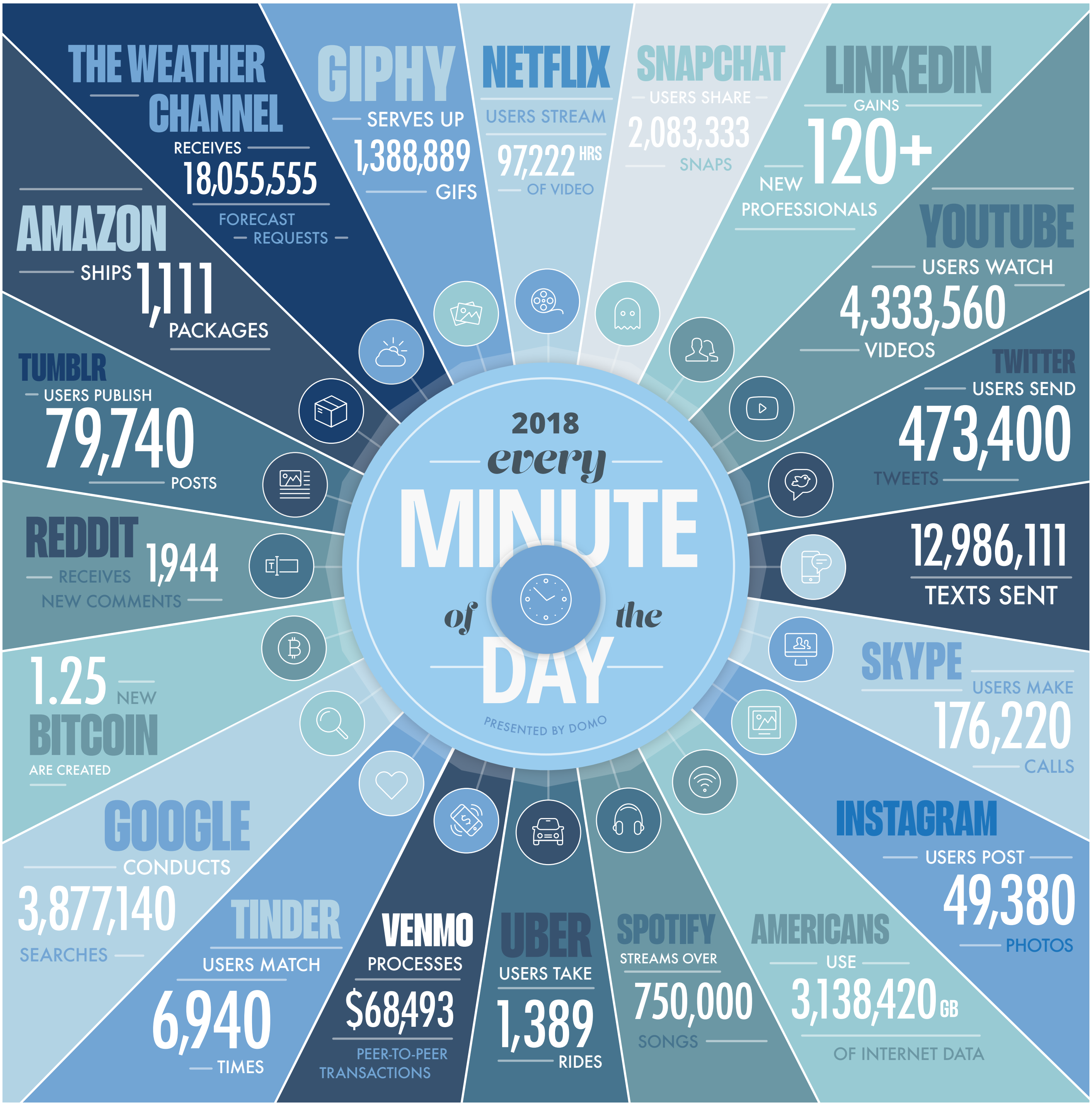
Key Contributor:

Qaisar Hasan
Portfolio Manager for the 1798 Q Strategy
Lombard Odier Investment Managers (LOIM)

For further information please contact:

Gene.getman@lombardodier.com

FIG. 1 SAMPLE OF THE VOLUMES OF DATA CREATED EVERY MINUTE[1]



[1] Image from Domo, "Data never sleeps 6.0," 5 June 2018. Data from: Statistica, LinkedIn, Internet Live Stats, Expanded Ramblings, Slash Film, RIAA, Business of Apps, International Telecommunications Union, International Data Corporation.

Overview

Technological advancements have become integral to humanity. The internet now affects every part of daily life. It has transformed how people receive and review information, keep in touch with friends and work as well as relax. Phones have advanced more in the last ten years than computers had in the decades before them. Most people could not have predicted how this new degree of connectivity would shape the world culturally. From the younger "screenagers" to nations' Presidents, the trend of constant sharing, posting, commenting and tweeting has been adopted by everyone seemingly overnight. For evidence, one can look at the last two Papal Inaugurations, just 8 years apart (Figure 2).

FIG. 2 ST. PETER'S SQUARE, INAUGURATION OF POPE BENEDICT XVI (2005) AND POPE FRANCIS (2013)

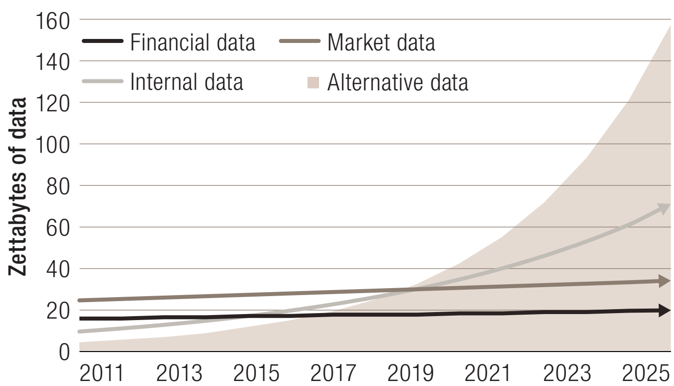
Source: Associated Press, Luca Bruno (2005), Michael Sohn (2013).

Since that photo in 2013, we have added nearly 1 billion more internet users to this planet. Even the Oxford English Dictionary got in on the fun, making "selfie" and 😂 their recent "word(s) of year."[2] We like to call this dramatic cultural shift "*the digitalization of everything we know*."

The general population largely sees all of the streaming, liking, posting and e-shopping as random data, but as analysts we study the powerful stories it can tell. The collection of digital footprints from a single individual may not be interesting, but when anonymized and amalgamated with millions of other people it allows us to draw connections and identify patterns. The volumes of data that humans, businesses, governments, machines and connected devices create every minute is more than people know what to do with. That in essence is Big Data. It is the incomprehensively vast amounts of real world commercial data that we create every day and the attempt of data scientists to make sense of it with the help of new technologies and statistical approaches.

The investment community has begun talking more and more about Big Data over the past few years as it realized what this information could teach us about business trends in near-real time, making us potentially better stock pickers. What's more, this new "alternative" form of data has just surpassed the available market and financial data that investment managers have based their investment insights on for decades… and it is just getting started (Figure 3).

FIG. 3 EXPONENTIAL GROWTH IN ALTERNATIVE DATA CREATION



Source: IDC's Data Age 2025 study, April 2017.

As Jim Simons once said in relation to data sources for investing, "everything is grist for the mill, weather, annual reports, quarterly reports. The historic data itself, volumes, you name it, we take in terabytes of data a day and we store it away, and massage it, and get it ready for analysis."[3] Naturally as the type of data and technology around processing it evolves, so must our investment approach. However, if data is gasoline and machine learning is the combustion engine, this new form of alternative data is crude oil. It needs to be thoroughly refined before it can even be considered fuel. In this piece we explain the unique differences of Big Data and the new challenges it presents for both quantitative and discretionary managers alike.

[2] Oxford University Press, Oxford English Dictionary word of the year in 2013 and 2015.

[3] Ted.com, Jim Simons: The mathematician who cracked Wall Street, 3 September 2015.

Equity investing is an extremely competitive landscape. There are more equity long/short hedge funds out there than stocks in the Russell 3000; roughly USD 1 trillion[4] of capital broadly trafficking in the same pools of intelligence for alpha. Traditional sources of information that fundamental analysts use such as sell-side research, investor conferences, management team interviews and regulatory filings have been the cornerstone of legacy investment programs for decades, but crowding and passive investing have made it more difficult to find a differentiated edge. Managers go to great lengths in pursuit of an informational edge: differentiated insights lead to more alpha, lower correlation to markets and peers and therefore better diversification for investor portfolios. It is therefore no surprise that we hear more hedge funds talking about purchasing Big Data or speaking about a more "quantamental" approach.

All that being, said Big Data is far from a panacea. People tend to talk about the success stories when it comes to Big Data, but there is often massive overfitting and live models are rarely correct. There are many misconceptions around the use of Big Data and skepticism is warranted. We argue that while one in five equity managers can claim to use alternative data as an investment consideration, the challenges of extracting the 2-3% of actual investment insights out of all the noise is an arduous task. The resources to purchase an expensive credit card panel does not guarantee the manager will get any return from it. Successfully extracting investment insights from Big Data is less about affording the golden ticket, and more about establishing a process that can consistently find needles in haystacks.

To understand and address these topics, we seek to answer the following questions:

I. What is Big Data?
- **a.** Concepts and definitions
- **b.** A recent phenomenon
- **c.** Types and sources of alternative data
- **d.** How is it measured
- **e.** Today's landscape

II. What can Big Data tell us?
- **a.** A new source of fundamental insights
- **b.** The process of harnessing data-driven insights
- **c.** Appreciating its scope and limitations

III. What are the unique challenges to successfully using Big Data?
- **a.** Barriers to entry & learning curve
- **b.** Evaluating data
- **c.** Refining data
- **d.** Aligning investment culture

IV. How should investors think about implementation and costs?
- **a.** Man versus Machine
- **b.** The resource misconception
- **c.** Maintaining a lead

V. What are some examples of applying Big Data?

Please note:
For the purposes of this piece, terminology like "our," "ours," "we," "us" relates to the views of one or more professionals within the 1798 Alternatives group of Lombard Odier Investment Managers ("LOIM"). Namely, the 1798 Q Strategy, which champions and deploys many of the views in this piece.

[4] Figures as of Q2 2018, ~USD 1.5 trillion across all equity HF. Barclays Strategic Consulting analysis based on HFR and HFI data, "Finding Alpha: Developments in the equity hedge fund landscape," September 2018.

What is Big Data?

a. Concepts and definitions

A little over a decade ago, data was limited to what people manually entered into computers. This concept changed as the internet shaped our culture and users began generating their own data through social media and content sharing. Accessibility and technological advancements further accelerated data integration and growth. The next major evolution came from developments in artificial intelligence, digital sensors, and connected devices, which allow machines to build links and record data on their own. Today, we create data in greater orders of magnitude than humanity ever thought possible. Nearly 5 billion gigabytes of data is generated **every day**[5] in real time across a vast range of mediums and sources. Nearly all of this data is different from the traditional structured datasets that investment analysts had become accustomed to and unique from the data that legacy investment programs are based on. We refer to all this new information as “alternative” data.

Alternative data: We define it as any dataset that does not meet the criteria for traditional financial data (e.g. income statements, balance sheets, press releases) or market data (e.g. pricing, volumes, factors). It is highly diverse and predominantly noise, but does offer a window into real-world commercial activity. While not all alternative data necessarily comes from Big Data (such as government statistics or trade data), nearly all of the growth in alternative data has come from Big Data sources. For the purpose of this piece, we refer to Big Data and alternative data synonymously.

Big Data: The literal definition of Big Data is a term applied to complex data sources, where the massive scale, frequency and abstract nature of the data is beyond the ability of traditional relational databases to capture, manage and analyze. More practically, we will reference the broader concept of Big Data, encompassing all of the breakthroughs and processes that allow us to make sense of unstructured data for the first time in history.

We think of Big Data as not only the data itself, but also as the set of skills and technologies that capture, store, normalize, analyze and visualize large variable collections of data in an attempt to gain unique investment insights about a company. The practical application of Big Data therefore also demands expertise in **Data Science & Data Analytics**. These terms both refer to advanced analytical techniques that require knowledge of data sourcing, data mining, data warehousing, programming algorithms, statistical modelling, machine learning, natural language processing and IT visualization tools. These skills are crucial to refining tens of millions of unstructured signals into actionable investment insights.

b. A recent phenomenon

In relation to machine learning and statistical methods that have existed for some time, the proliferation of Big Data is a relatively recent event, developing materially over the past few years. We believe the newfound capability to extract value from alternative data is a function of four key trends:

i. Computational speed and decreasing costs in cloud storage

Moore’s Law is the observation made by Intel co-founder Gordon Moore that the complexity and computing power of hardware doubles every 18 – 24 months, while costs are halved. This exponential growth eventually gave us the technological innovations in computing speed, data storage and network infrastructure necessary to perform advanced data science.

Over the last 50 years, Moore’s Law has solidified itself as the golden rule for the technology industry. Another significant, but much more recent development is the advent of cloud computing and storage. This area of technology has really taken off over the last five years and has been revolutionary for the ease of access to affordable computing capacity and storage. Its anecdotal cousin is “*Bezos law,”* which states, “over the history of the cloud, a unit of computing power price is reduced by 50% approximately every three years.”[6] While that has not exactly played out, cloud-computing costs have still come down 76% in the last 10 years.[7]

ii. Digitization of consumer lifestyles and business processes

Technological advancements have paved the way for globalization and changed how people communicate with each other, transact and interact with businesses.

Access to the Internet continues to shape the digitalization of consumer lifestyles. In the last five years, Internet access has gone up by 1.3 billion people, with usage now reaching 47% of the world’s population.[8] While the pace of this trend is impressive, one of the key aspects that has shaped our cultures is the medium in which we connect. As shown in Figure 4, global internet traffic will continue to grow 30% annually, increasing threefold in just five years. More importantly, nearly all this growth will come from smartphones, which are projected to account for 50% of all internet traffic by 2022. Smartphones are filled with a myriad of sensors that collect data on a very granular level, generating a wealth of information. Our locations, spending habits and online behaviors have been datafied, anonymized and indexed for analysis.

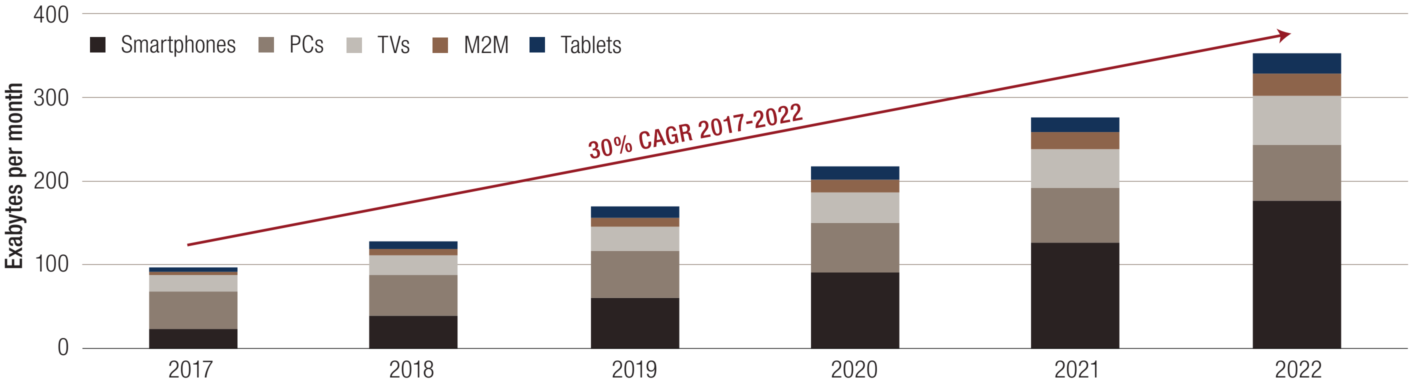
5 Cisco VNI Global IP Traffic Forecast, 2017-2022, trend extrapolated for 2018 snapshot.

6 GigaOm post by Greg O’Conner, former CEO of AppZero, 2014.

7 AWS Internet archive data from m1.large unit in 2008 to m5.large unit in January 2018.

8 Domo, “Data never sleeps 6.0,” 5 June 2018, Live internet stats.

FIG. 4 GLOBAL INTERNET PROVIDER TRAFFIC BY DEVICE



Source: Cisco VNI Global IP Traffic Forecast, 2017-2022. Other device types represent less than 0.2%.

Businesses are leveraging advanced logistics and digital technologies to better connect with customers and improve operating efficiency. They are adopting initiatives around e-commerce, digital payments, supply chain management, direct-to consumer strategies and data logistics to improve market share and margins. Similar to how digitally progressive businesses are using technology to better understand their customers, savvy investors are embracing the wealth of information created from these practices to more accurately value those businesses.

iii. Emergence of business intelligence tools to visualize large sums of data

More data does not necessarily mean better insights. On the contrary it often leads to more confusion, contradiction and stress without the proper visualization tools. One of the best examples of this is the healthcare industry, where new artificial intelligence (AI) tools provide doctors with data reports to help diagnose patients more efficiently. While this would seem like a great breakthrough, interviews with doctors show the opposite, describing it as a "massive monster of incomprehensibility" as they hunt through immense reports for relevant details. Despite good intentions, studies show this has actually led to higher levels of burnout and depression among clinicians.[9]

For this reason, a vital development area in recent years has been the emergence of business intelligence software and applications, which sort, summarize and present terabytes of data in a concise format. In order to derive actionable investment insights while the data is still relevant, users need tools that can help them to digest and present information in a way that is seamless and natural to our decision-making processes. Advancements in data analysis and affordable access to external vendors have recently paved the way for the business intelligence and visualization software necessary to make Big Data more practical. Some examples of providers include QlikView, Tableau, and Splunk.

iv. Ecosystem of Big Data intermediaries

Much of the growth of the Big Data ecosystem was actually born out of the exhaust from traditional commercial intent. Vendors began by selling analytics directly to businesses as ways to enhance organizational return (e.g. hotel statistics to lodging and leisure companies, ad spend data for the marketing industry, image recognition and diagnostic databases within the healthcare industry). These vendors eventually realized that this data might offer commercial insights for investing and began offering datasets to buy-side firms. That said, the total buy-side firms spend on alternative data only accounts for an estimated 1.3%[10] of the USD 32 billion global Big Data market.[11] This market and ecosystem is expected to continue to grow at an annual rate of nearly 20%, reaching USD 156.7 billion by 2026.

The growth of this buy-side focused vendor and intermediary ecosystem has provided new and affordable ways to leverage interesting data sets and specialized expertise. Vendors can sell both unstructured datasets, as well as semi-structured data, where their data-science teams have already curated the data to some degree by cleaning, interpreting and presenting the data. Some examples of Big Data intermediaries include 7Park, 1010data, and Earnest.

c. Types and sources of alternative data

Unlike traditional sources of investment insights that come from pricing/tick data or company filings, alternative data comes from a much wider range of sources.

Some of these sources, like government statistics, may be more in the realm of "small data" where the information tends to be more normalized, regular in frequency (e.g. monthly or quarterly) and offer much more history. The vast majority of growth in alternative data, however, has come from Big Data. This includes a broad

9 "Why doctors hate their computers," Atul Gawande, The New Yorker, 12 November 2018.

10 AlternativeData.org, 2017 estimated USD 400 million buy-side annual spend on alternative data.

11 Big Data – Global Market Outlook (2017-2026), Statistics MRC Pvt Ltd. October 2018.

range of sources such as geolocation, web traffic, video/audio, satellite imaging, credit card receipts, applications, web searches and social media.

Data from these sources is **Unstructured**, i.e., information that either does not have a pre-defined data model or is not organized in a pre-defined manner. Each type is independently gathered, anonymized and consolidated into large datasets, which data scientists clean and structure in an attempt to spot trends. Some of these sources are free and can be mined internally, but most datasets are purchased from vendors as either raw unstructured data or semi-structured data. Different data types and vendor approaches result in a wide range of frequencies (some are daily, others may be hourly, weekly or monthly) and in most cases lack more than 3-4 years of history. Combining multiple sources of data can provide a mosaic of information about a company, but normalizing this broad range of outputs into actionable insights is a demanding task.

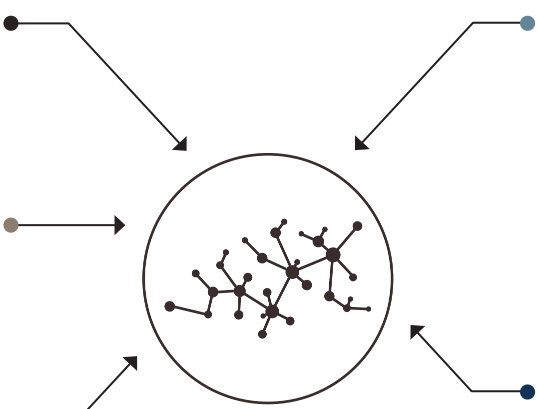
We tend to focus on five broad categories of alternative data shown below in Figure 5. This list is not all encompassing. While there are a number of other alternative data sources such as social media, we generally find them to be less reliable or relevant for our purposes.

FIG. 5 SOURCES OF ALTERNATIVE DATA

Consumer spending
- Credit cards
- Emails
- Point of sale terminals

Online behavior
- Web traffic
- Web searches
- Web scraping
- App downloads
- Clickstream
- Cookies

Geolocation
- Satellites
- Drones
- Smartphones



Government statistics
- Trade data
- DMV registrations
- Building permits
- Personal bankruptcy filings
- Local number portability
- Bill of lading

Contributory databases
- Point of sale
- Hardware distribution
- Hotel stays
- Media buys
- Flight bookings
- Drug prescriptions

d. How is it measured:

Big Data is generally assessed based on five characteristics:

Volume
The massive scale of the data that is being created and stored daily

Variety
The different sources and formats of the data being created

Velocity
The ability to keep up with the speed at which data is being generated

Veracity
The uncertainty of data, including relevance, reliability of the source, unintended biases, or general accuracy

Value
The costs and applicability of the data for deriving insights

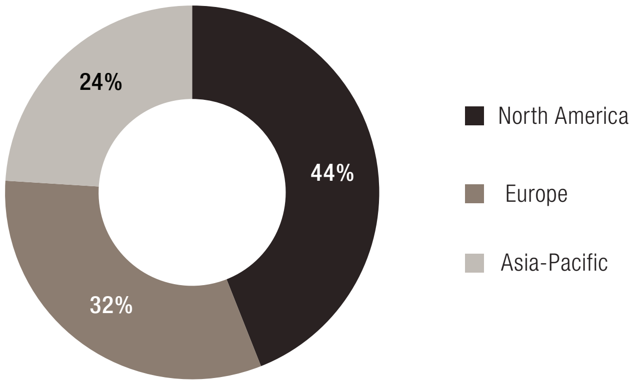
e. Today's landscape:

Ecosystem: The Big Data ecosystem supporting the buy-side has exploded within the past few years in terms of both growth and awareness. We now have Big Data conferences, intermediaries, and many more vendors to choose from. According to Neudata, an advisor to investment managers on alternative data, the number of vendors in their index has risen from 172 to 468 in the last year alone.[12]

Sector overview: It is no surprise that the consumer space (discretionary, services, and staples) offers the largest breadth and available history of alternative data. Analysis of Big Data began with the goal of better understanding consumer spending habits and online behavior to improve corporate marketing. Alternative insights help detail key revenue trends within retail, grocers, restaurants, hotels, leisure and entertainment. The next largest segment is TMT, and particularly internet, software, and hardware companies. Industrials has been slower to adopt data analytics outside of airlines and auto segments, however technological initiatives to manufacturing, shipping, and supply chain management have allowed for robust signals more recently. While other sectors are clearly investing in data analytics and infrastructure, our approach has been to focus on the availability and quality of data insights. As sectors like healthcare continue to invest heavily in data and logistics, we expect to expand our investment universe opportunistically into new sectors and regions over time.

Geographic overview: Higher rates of internet access and adoption of analytics has contributed to North America's domination from a geographical standpoint. Developed Europe has historically been the second largest region, but has been lagging behind relative to the growth coming from Asia-Pacific and China in particular. IP traffic in APAC has been growing at an annual rate of 32%, compared to 22% for the developed world.[13] While aggressive Chinese investing into technological infrastructure has fueled growth in data generation and analytics, other factors like dovish privacy considerations and rising internet penetration across the region have also played large roles. We are already seeing material diversification in the regional focus of the data vendors (Figure 6) and expect this trend to continue over time.

FIG. 6 REGIONAL DISTRIBUTION OF DATASETS



Source: Neudata, Industry trends presentation, November 2018. Figures as of 12 October 2018, exclude all other regions and datasets defined by Neudata as global.

[12] Neudata, Industry trends presentation, November 2018. Figures as of 12 October 2018.

[13] Cisco VNI Global IP Traffic Forecast, 2017-2022.

What can alternative data tell us?

> More data doesn't just let us see more of the same thing we were looking at. More data allows us to see new, it allows us to see better, it allows us to see different.[14]
>
> *Kenneth Cukier*

a. A new source of fundamental insights

It seems contemporary perception of Big Data is that of a tool – an overlay or product that can be purchased and implemented into legacy investment frameworks – that turns discretionary managers into "quantamental" or adds to the vast toolkit of quantitative managers. We believe this is a common misconception and prefer to think of Big Data instead as a new wave of thought that is worthy of a new investment framework. While the investment philosophy behind alternative data is inherently fundamental, the foundation is quantitative by necessity. That said, its unique form and challenges make it difficult to bucket as solely discretionary or quantitative.

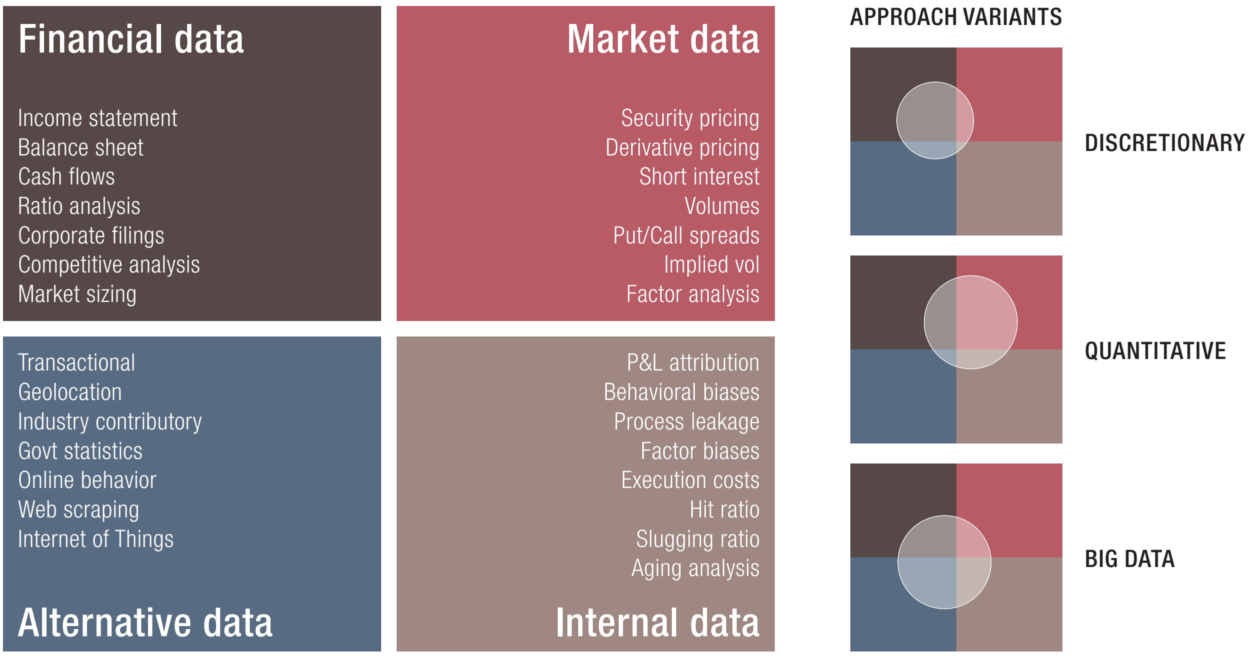
Waves of thought in investing:

i. **Discretionary** – Very simplistically, stocks follow earnings and earnings follow revenues. The idea being companies that continue to beat on revenues tend to relate to companies that see share price appreciation over time. Discretionary managers develop predictive models using **Financial Data** (Figure 7), as well as adjustments based on intuition from reading sell-side research, attending conferences and conducting management team interviews.

ii. **Quantitative** – Predicts security prices, using mathematical and statistical modelling to identify links between pricing, valuation, and other **Market Data** over time. It is based heavily on clean and structured data, and the underlying algorithms can explain where share prices may go, but not why.

iii. **Big Data** – Tracks real-time operational and commercial activity, to make predictive links to company-specific business trends. The data tracks earnings and inherently speaks to fundamental trends, but the skills necessary to process it is a subset of machine learning strategies and therefore inherently quantitative. It is based on the analysis of new, messy **Alternative Data** and what it can tell us about being a better stock picker. Using alternative data without understanding the fundamental context that it represents is very challenging.

FIG. 7 THE FOUR DATA QUADRANTS OF ANY INVESTMENT FRAMEWORK

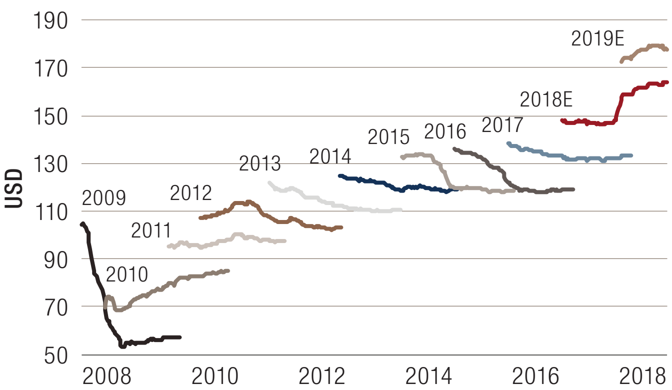


[14] Kenneth Cukier: Big Data is better data, TED Talk, 23 September 2014.

Fundamentally, we ask similar questions

How many iPhones will Apple end up selling? How many subscribers will Netflix end up having? How many jackets will Macy's sell? We are fundamental investors at heart and believe a company's value is ultimately a reflection of its long-term earnings capacity. Differentiated opinions form the foundations of how investors will long or short a pool of companies. Nevertheless, empirical evidence shows humans are not good predictors – by nature we exhibit strong biases that reflect our personal experiences, prior successes (or failures) and anecdotal data points – so anticipating where the world, let alone single companies are headed is a steep challenge. For example, we can look at just how much annual consensus EPS for the S&P 500 shifts throughout each year (Figure 8).

FIG. 8 CONSENSUS FORECAST S&P 500 EPS FOR EACH YEAR



Source: Morgan Stanley Research. Figures as of February 2019.

So how does a fundamental investor go about speculating the future without being run over? We believe the answer lies in building an objective, systematic and scientific process underpinned by evidence, which in our case is data. When implemented correctly, such a process can provide very precise, real-time insights into near-term earnings inflections while keeping personal biases at bay.

Qaisar Hasan, Portfolio Manager for the 1798 Q Strategy had his first foray into Big Data back in 2011, when he developed a model using google trends data to predict quarterly Netflix subscribers. Today's models have evolved to include data from app usage, web traffic, credit card transactions, and IP information. The added complexity and speed of the data requires a process built specifically around extracting precise insights from alternative data.

b. The process of harnessing data-driven insights

Alternative data is profoundly different from traditional data sources. The diversity and nuances of the data mean that you cannot simply back test all of the data available and select the best fit. A process to filter and prioritize Big Data efforts is necessary for success.

While there are numerous ways to design an investment process around alternative data, they all try to answer two key questions (1) How do we use all of the data coming in rapidly to better understand the companies we invest in? (2) How does alternative data help us draw conclusions more accurately?

Our investment process at 1798 Q[15] is a bottom-up one, focusing on what data points are telling us about current business trends at a very specific company and key-performance-indicator (KPI) level (Figure 9). It is a blend of machine learning and artificial intelligence systems designed to curate the massive quantities of data, as well as human ingenuity. Creative thinking is vital throughout the process to design the logic behind the algorithms and extract value from the data. Broadly speaking, the process involves (i) evaluating data, including sourcing and back testing, (ii) data science, tagging and mapping data with the help of proprietary algorithms and (iii) contextualizing data into an investment thesis.

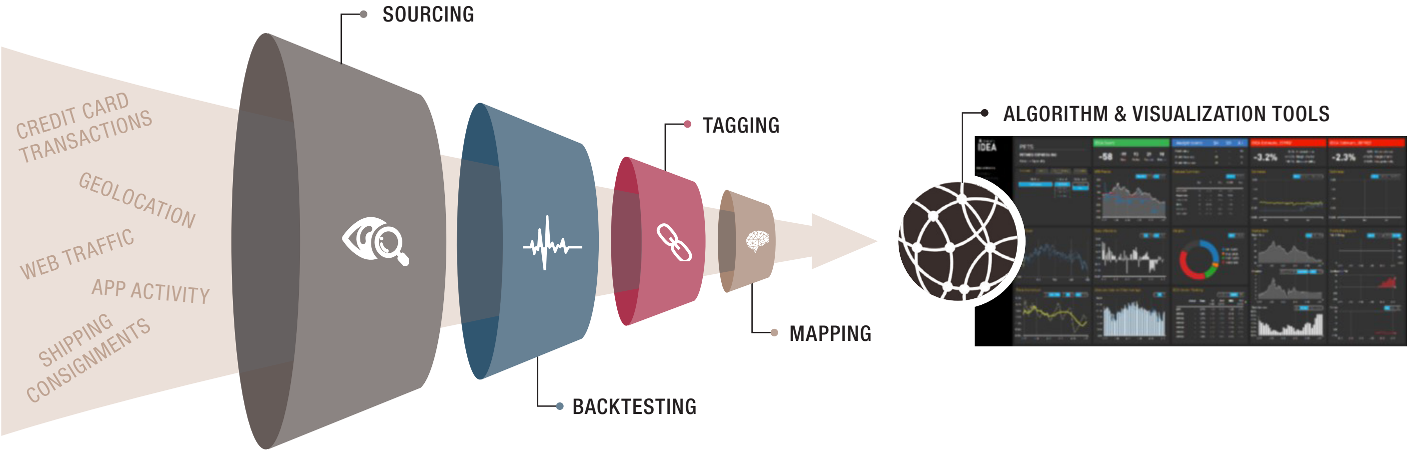
Sourcing

Sourcing is finding vendors that have something interesting to show you. In the early days of the industry this meant going to trade conferences for various industries (e.g. automotive, chemicals) and asking data vendors what do they do in the industry? What role do they play? What resell rights do they have for their data? The goal is to understand what this data could tell you: what does it capture, what verticals does it cover?

With the maturation of the vendor ecosystem, this process has been somewhat streamlined. There are now thousands of datasets aimed at the buy-side, but access to more data for analysis does not mean you gain more insight. On the contrary, today's environment is more like drinking from a fire hose. Properly evaluating, tagging and mapping a single dataset can take from weeks to months without the proper systems in place. Therefore before ever applying data science, it is crucial to develop filters to maximize time value from a commercial point of view and short list data sets based on fundamental need. The first question we need to answer is do we want to do more work with the data?

[15] The portfolio information provided in this document is for illustrative purposes only and does not purport to be a recommendation of an investment in, or a comprehensive statement of all of the factors or considerations which may be relevant to an investment in, the referenced securities.

FIG. 9 SAMPLE INVESTMENT PROCESS BUILT SPECIFICALLY AROUND ALTERNATIVE DATA



Source: 1798 Q "IDEA" Research Platform.

The key is not assessing whether the data can make us money, but whether it can help us predict company fundamentals more accurately. This qualitative process involves asking more questions such as:

- Does this data make us smarter?
- Does this data match up against problems we are trying to solve for?
- What could this data mean for company revenues and business models?
- Do the insights exceed the cost of acquiring and managing the data?

Sourcing data successfully requires a fundamental knowledge of what moves stocks as well as thinking creatively about what specific types of data could tell us about different types of companies:

- A geolocation dataset provides an idea of foot traffic within physical locations that are designated to certain businesses. GPS chips in phones ping locations to servers roughly every 5 minutes, providing a good idea of consumer traffic to a specific store or location. This may offer a small glimpse into some companies, but for businesses like theme parks for example, it can provide a strong indicator of daily revenues on a very granular level.
- Satellite images can be used to gauge retail sales and customer interest by observing the number of cars in parking lots. It can also be useful for detecting real-time levels of oil inventories, crop output, manufacturing activity, etc.
- For businesses with heavy e-commerce or direct-to-consumer models, transactional data such as credit card receipts can tell us a great deal about their sales.
- Smartphone application data, like unique downloads and usage statistics, can tell us a great deal about new customers and their engagement patterns. This can have varying applications, from understanding the number of Spotify subscribers to new Tesla owners.

Backtesting

Based on experience and whether the data can be applied to fundamental models, we typically reject 70-80% of datasets before more rigorous analysis even takes place. Once the more promising datasets are narrowed down, the data needs to be run through systems to see where it applies and if there is value in implementing it. As we will discuss in more detail in the 'challenges' section, the reality is almost 95% of data is garbage, apt to giving users false or inconsistent signals. This is why robust evaluation criteria and understanding of data biases and limitations is vital to success. Backtesting is both fundamental and quantitative and the process is not perfectly linear. We evaluate roughly 100 new datasets a year to narrow down the 10-15 datasets we may want to work with.

Tagging

Tagging is the process that effectively converts unstructured datasets into structured indices that are ready for analysis. This is mostly systematic, using machine learning and AI tools like simple neural networks to normalize, scrub and refine datasets into series that can be analyzed for trends. It is an extremely time consuming exercise to develop the proper infrastructure and managers have to make the decision as to whether they will focus the resources on tagging the data in house, or leveraging the vendor ecosystem.

Mapping

Once the data is structured into a clean index it can be mapped to specific companies. We identify key metrics that will drive company earnings over a 3-6 month horizon and try to find data sources that will give us transparency. There is a great deal of taxonomy involved as the data does not just tell us whether a company like Amazon may beat or miss earnings. An index typically speaks to very specific metrics, often at a brand level. The mapping phase is where the data is categorized by relevance to a particular segment or financial metric of a company (e.g. online revenues of Zappos.com, a subsidiary of Amazon). Furthermore, different brands may have different accounting standards, transaction dates and other nuances that must be considered. For a company like Amazon, we use a mosaic of information from different alternative data sources to paint a robust picture. We may use credit card transaction data to look at Amazon.com revenues, geolocation data for Whole Foods Market, and reverse IP lookups to get a read on the Amazon Web Services business.

Visualization

In order to translate data signals into actionable insights, it is important that we systematize and contextualize the data with front-end visualization tools. These are essential for a fundamental analyst to extract and understand the data quickly enough to take action while signals are still topical. For this reason, it is important to design visualization tools that are intuitive to our fundamental process and portfolio management practices. Our “IDEA” research platform has been custom built for extracting alpha from alternative data down to the issuer level.

c. Appreciating its scope and limitations

A rather obvious drawback of data is that it is inherently backward looking. One needs to embrace the limited scope of data relevancy over time and build an investment process that aligns idea turnover with the strengths and weaknesses of Big Data.

If modeled well, recent data trends can be extrapolated into the near future. Nevertheless, predictive power diminishes the further out you project, just as a flashlight may allow you to see 10-20 feet in front, but not a mile ahead. This is where the concept of **incrementalism** comes in. Pieced together in one-month rolling cycles for example, we believe data can be a strong indicator of the future inflections.

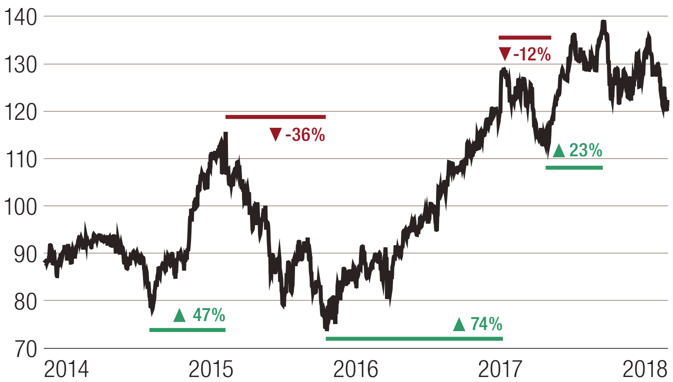
To illustrate this point see Figure 10. It represents the aggregated share price performance of eight of the largest EU-based Industrials companies. This group is highly complex, with multiple product segments servicing clients across the globe, and frequent M&A transactions influencing their earnings makeup. Industrials by nature are also quite cyclical, as reflected by the large performance moves experienced by the group every year.

In Figure 11, we overlaid the share price index for this group with a real-time projected revenue growth index for the same set of underlying companies using multiple data signals. This point-in-time revenue growth index includes several macro- as well as company specific alternative data inputs, and represents the projected revenue growth for each quarter, often weeks prior to the companies releasing earnings.

A few striking things emerge from this chart. First, alternate data has been able to catch nearly all major earnings inflections in real-time, despite the complexity of the underlying businesses. Second, rather than taking a sanguine long term view of an inherently cyclical space, the market prices these companies based on quarter-to-quarter fluctuations in their short-term earnings prospects.

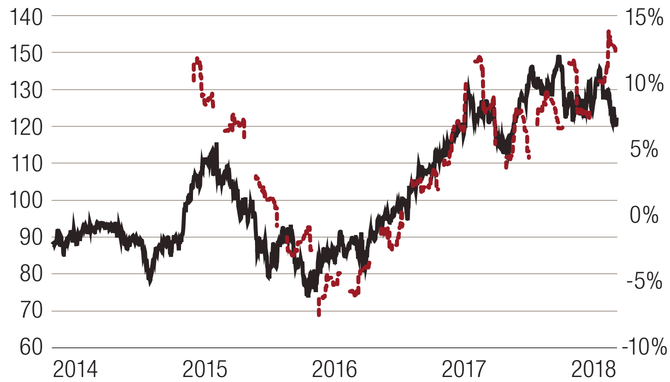
A prescient fundamental investor taking a bullish long-term view on EU industrials back in 2014 would have reaped a healthy 9% return through the end of 2017. On the other hand, an incrementalist approach, using data driven insights to trade periodic inflections in earnings, would have been able to capture the numerous bull/bear cycles experienced during this four year period, with considerably more confidence and visibility.

FIG. 10 EU INDUSTRIALS SHARE PRICE INDEX



Source: Bloomberg. Simple average share price of eight large EU Industrial companies.

FIG. 11 PRICE INDEX VERSUS REVENUE GROWTH PROJECTIONS



Source: Bloomberg. Simple average share price of eight large EU Industrial companies. Proprietary LOIM data analysis using the 1798 Q “IDEA” Research Platform.

What are the unique challenges to successfully using Big Data?

a. Barriers to entry and learning curve

There are many pitfalls to using Big Data, and a lot of the cynicism we hear about using it successfully is warranted. Data is not a cure-all, and 95% of it is often incorrect or misleading when taken at face value. The key is developing the processes and experience necessary to filter out the noise and extract the 2-3% of data that actually provides meaningful insights. While it is entirely possible, the barriers to using Big Data successfully are quite steep.

Building out a robust data science platform and proper visualization tools requires considerable capital investments and roughly 12-18 months of development before investors can reap any benefits. Costs can run into the multiple millions of dollars just to maintain the talent and datasets necessary to continue to run the operation. Even when managers commit to the buildout and costs, they often find that they cannot extract enough consistent alpha to justify the upkeep due to a mismatch to their investment culture/framework (discussed later).

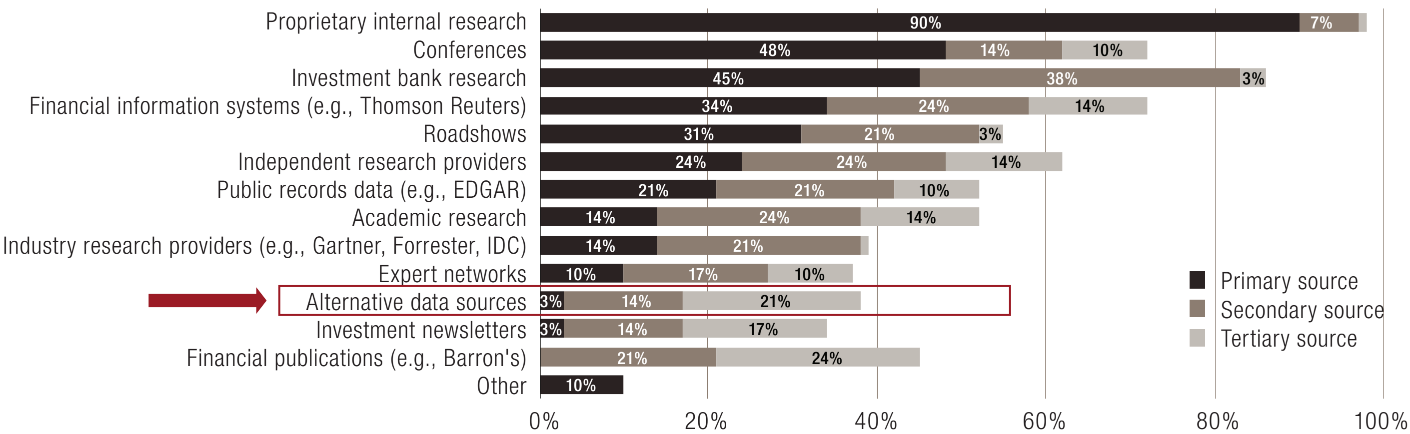
As highlighted below in the recent survey conducted by Greenwich Associates (Figure 12), alternative data is at the bottom of the list when it comes to sources of primary investment research, coming in at just 3% of managers. That said, 14% of managers claim it is a secondary source of research, and 1 in 5 managers claim it is a tertiary source of research, the second highest of any category. There is a great deal of pressure for hedge funds to adapt in search of alpha, but we find that the hurdles to using Big Data are greatly underappreciated. People only talk about the success stories of using Big Data, but many times managers purchase alternative data sets only to find that there is massive overfitting or predictive bias, and the program falls apart when implemented in live models.

New information like credit card activity, online browsing history and foot traffic may uncover additional clues about current trends, but in truth, there is no assurance of additional alpha. A disciplined framework built around objective tools and procedures is essential to evaluate datasets and understand the risks. The key to success is balancing technical proficiency with good judgement. There is no short cut when it comes to developing good judgement around alternative data; breaking new ground comes with a steep learning curve.

Qaisar Hasan began his career as a fundamental analyst and began applying data science to alternative data in 2011. Whereas street estimates generally have an error rate of about 2.5% around predicting company revenue, his initial error rate was as high as 4.5%. Over the course of seven years, this percentage has dropped down to 1.2% today,[16] with a path below 1% in near sight. While the data ecosystem and increasing scale have allowed for robust insights by collecting a wide variety of different data, both the machine and human elements of the process have an unavoidable learning curve.

Within the proprietary systems, the algorithm's **feedback loop** feature learns from successes and failures over years of earnings reactions to improve future predictions. Just as important is the learning curve of the architect behind those systems. Aside from data application ingenuity, this includes experience in monitoring and adjusting the system's recommendations based on numerous risks and considerations that machines cannot adequately capture (market conditions, liquidity concerns, event path).

FIG. 12 HEDGE FUND SURVEY OF INVESTMENT RESEARCH SOURCES



Source: Greenwich Associates 2018 Future of Investment Research Study, 28 June 2018. Survey based on 30 respondents.

[16] Past performance is not a guarantee of future results. The portfolio information provided in this document is for illustrative purposes only and does not purport to be a recommendation of an investment in, or a comprehensive statement of all of the factors or considerations which may be relevant to an investment in, the referenced securities.

b. Evaluating data

Perhaps the most critical distinction among investment managers using Big Data is their ability to gauge the accuracy and value of a broad range of datasets. The relationship between the amount of data consumed and the amount of insights attained is not linear. While there is greater access and breadth of data than ever before, the growth of the ecosystem also brings forth more noise and false signals. Investment managers must build a process that emphasizes evaluation tools in order to maximize return on investment capital (ROIC) and time (ROIT) and optimize their net returns.

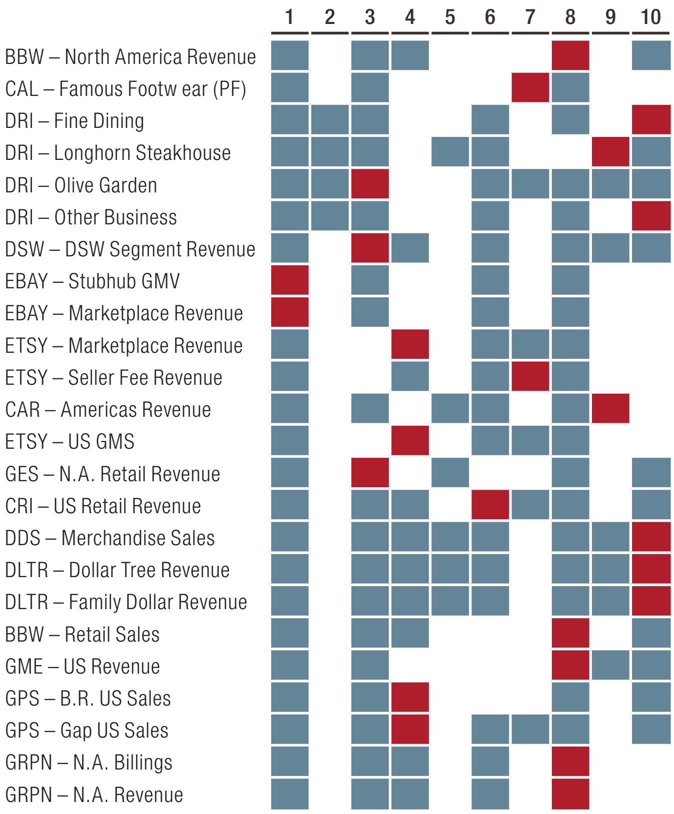
First off, not all types of alternative data can be trusted. In his book Everybody Lies, Seth Stephens-Davidowitz conducted a comparison study of two publications, *The Atlantic* (a high brow policy publication) and *The National Enquirer* (A decidedly low brow gossip rag). Seth compared the number of Facebook "likes," article clicks and monthly readership between the two publications and his results quickly revealed that even though we "liked" Atlantic articles 24 times more than the National Enquirer, actual circulation data indicates nearly identical readership bases. To no surprise, the author was able to identify that as a general public, we want to "appear smart and cultured" yet in reality all we want is lowbrow, easy to consume, entertainment and gossip.[17] This helps to highlight why sources like social media and survey data must be taken with a grain of salt. Even anonymous surveys can carry similar biases; hence the surprise results in the 2016 US Presidential election and the earlier Brexit referendum relative to polling data. "Many people underreport embarrassing behaviors and thoughts on surveys. They want to look good, even though most surveys are anonymous. This is called social desirability bias."[18]

Even within more trusted types of data, investment managers must be able to appraise the incremental value between various vendors. As an example, a dataset that gives transparency into the activations of new iPhones may speak to 70-80% of Apple Inc.'s sales, but the onus is on the investment manager to select one of the 15-20 vendors that offer iPhone activation data. While one would think the differences among them would be negligible, the dispersion within this relatively straightforward dataset is shockingly large. Alternative data is not binary and vendor interpretations vary based on a multitude of assumptions, limitations and biases. Therefore, even structured data cannot be taken at face value. Vendors do not discriminate or discern whether the data has any investment value. They are simply focused on monetizing their data sets and the buy-side is a small fraction of the Big Data market.

The burden of responsibility for proper evaluation falls on the investment manager and it does not stop at vendor selection. Taking this a step further, we believe real competitive edge within this space is tied to the ability to discern the quality and potential return of a new dataset at an individual stock and KPI level. For a real-world example, we can look at Figure 13, which shows 24 different KPIs that we have identified across 14 companies. Across these names, we provide a sample of 10 different datasets, highlighting the most historically accurate dataset for each metric in red, based on our proprietary analysis.

What becomes clear is that no one dataset holds a monopoly over accuracy and this holds true with much larger sample sets. Some datasets may span hundreds of companies, but they also incur considerable costs to acquire and refine and even then may only lead to accurate conclusions on a few companies. At the same

FIG. 13 ACCURACY VERSUS COVERAGE OF SAMPLE DATASETS 1-10



Source: Proprietary LOIM data analysis using the 1798 Q "IDEA" Research Platform.

[17] Everybody lies article, Countertrend Digital, written by Daniel McCarry, 2 November 2017.

[18] Everybody lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are, Seth Stephens-Davidowitz, 9 May 2017.

time, some of the best datasets are often overlooked because they only cover a handful of names. In order to maximize ROIC when using alternative data, it is necessary to balance the incremental costs and insights that a new dataset may carry, relative to both the broader investment universe and existing datasets at a micro enough level.

Furthermore, it is essential to appreciate the costs associated with time-spend. Structuring datasets in house incurs meaningful human capital to clean. Even within pre-structured vendor solutions, it can take over a month to properly evaluate and integrate a single dataset. With no set standards between vendors in terms of data format, frequency, or history, even deciding between two vendors selling the same data is often not apples to apples. This challenge grows exponentially when evaluating 500+ vendors offering different types of data. From our perspective, we believe future success is a function of efficiency and scalability, which is why we place so much emphasis on custom-built systematic processes that allow us to screen 100+ vendors a year and onboard new datasets in a fraction of the time.

c. Refining data

Alternative data is generated by people and therefore requires context. It is messy, complex and constantly evolving so there are a number of decisions that need to be made from a normalizing and modelling standpoint, which can lead to very different interpretations. Even a vendor scrubbed panel will have many permutations of the same dataset and 30 different investment managers can derive 30 different conclusions.

Normalizing data

There are a multitude of decisions that must be made to address compatibility issues. Credit card indices for example may not always include merchant name or other important details, while geolocation indices are linked to specific cell phone applications. What happens when these apps get shut down or new ones get added? How are the different app user demographics considered to avoid biased models?

Treatment of errors and outliers

Vendors can make mistakes around things like currency. Investment managers must figure out how to validate data and treat outlier transactions on a company-by-company basis. A USD 1,000 purchase at Burberry may be typical, but how does one treat a USD 1,000 purchase at McDonalds?

Reflecting change

Underlying constituents go in and out of data indices periodically and models must evolve as datasets evolve. Companies themselves are also living organisms. They can make several changes ranging from accounting policies, transformative acquisitions, to moving into new product segments or geographies.

Given the sheer volumes of data and rate of evolution, it is crucial to develop a set of systematic rules and procedures to consider various issues in real-time. Designing these systems with scalability in mind is also key, as they are bogged down by all the different schema and nuances of every dataset very quickly.

d. Aligning investment culture

The least obvious headwind to Big Data implementation by legacy investment managers is their culture. The unique challenges and restrictions around Big Data require managers to develop an investment process built around extracting alpha from data. By nature however, humans are slow to evolve and change the ways of thought that have served them well for decades. The reality is one cannot simply read a Big Data analysis report and act, like how you would with a quantitative risk report. Big data is not a plug-in, tool, or overlay. It presents meaningful hurdles to discretionary and quantitative legacy programs alike.

Discretionary managers

Discretionary CIOs and Portfolio Managers have grown up through the ranks, investing in a certain way based on bottom-up fundamental analysis and gut instinct. Since this method has had nothing to do with data for decades, established investment managers often resist the push to drastically change the way in which they research ideas. There is often limited buy-in from legacy managers outside of referencing alternative data as a research consideration in marketing materials. The reality is that after some time, we see the same patterns of high turnover in data science teams, and PMs claiming the experience was a costly distraction. There are several reasons behind this:

- **Siloed roles:** PMs develop a thesis on where a company's revenues are going, and can then compare their models with alternative data to either confirm or refute their preexisting hypothesis. Data is usually a secondary input and PMs lack the time and technical acumen to evaluate and source the data themselves. Oftentimes they hire a group of token data scientists, give them a budget and resources and have them develop their own models. This group is generally a middle/back-office function, lacking any involvement in investment committee meetings or having a say in the investment process. This partition of roles creates friction. When the data refutes the manager's intuition, whom do they listen to? What happens to the team dynamic when data inevitably results in losses on a particular trade? This siloed structure also limits the ability of the data scientists to think like fundamental portfolio managers when speaking to vendors or designing the visualization tools.
- **Investment horizon:** As discussed in the 'Incrementalism' section earlier, is the legacy investment process nimble enough with respect to idea turnover to extract alpha while the data is still pertinent? The answer in most cases is no. Discretionary managers are accustomed to building longer-term DCF models, building a thesis that generally spans years, not months. The average L/S managers holding period on longs tends to be north of 12 months, while shorts are typically in the book for 9-14 months.[19] Given alternative data speaks to real-time trends lasting only a few weeks, what does a manager do with their conviction when the short-term data contradicts their long-term hypothesis?
- **Turnover:** The lack of a uniform culture generally leads to higher team turnover. Funds need to create organizational habits and communication channels that value all the roles equally. Friction between the investment and data science teams, coupled with high demand for qualified data scientists often results in high turnover and a loss of intellectual property.

Quantitative managers

Big Data would seem like a natural fit for quantitative strategies. Quants are used to working with large quantities of data and designing algorithms to sift through it all to spot trends. The reality is alternative data is completely different from what they are used to working with, leading many of the large quant funds to avoid it altogether

- **Unstructured fundamental data:** Quantitative strategies find the link between valuation multiples or other factors and share price data over time, largely by using clean market data. Fully systematic processes find alternative data to be too unruly due to its unstructured nature. Additionally, alternative data typically speaks to earnings and not to share price, so structuring it for use in legacy quantitative models requires a very labor intensive process using a different school of thought.
- **Limited history:** Quantitative strategies are used to working with very long historical datasets, which they backtest to identify share price patterns over the course of multiple market cycles. Alternative data on the other hand will often have only 4-5 years of history at most. A dataset with 4 years of history for example, will only have 16 quarterly earnings observations to backtest against. This means alternative data is extremely susceptible to overfitting, lending itself to hundreds of signals on a single stock, but blowing up in a live test model.
- **Limited scope:** Quantitative strategies generally spot patterns across thousands of stocks, diversifying smaller bets across a huge number of positions. Alternative datasets on the other hand tend to be very specific, often only speaking to 50-100 companies, with the best ones focusing on only a handful of names. This lack of scope makes the long evaluation process of most alternative datasets just not worthwhile given the size and construction of their legacy portfolios.

[19] Barclays Strategic Consulting survey, Finding Alpha: Developments in the equity hedge fund landscape, September 2018.

How should investors think about implementation and costs?

While many are aware of the groundbreaking defeat of World Chess Champion Garry Kasparov by IBM's AI machine, *Deep Blue*, in 1997, a less known story is the outcome of the first ever computer-assisted "freestyle" chess tournament held in 2005.

After his loss, Garry Kasparov proposed a first of its kind online chess tournament with two general styles of teams: (1) fully AI or (2) cyborgs, where the world's greatest chess players teamed up with AI machines. What was clear was that the latter group had a clear advantage, beating out hundreds of fully AI competitors. The final four contestants were all hybrid teams. Three of these teams were Russian Grandmasters running AI models on industrial strength computers, but the fourth team, under the alias of "ZachS," was a complete mystery to the chess world. Rumor had it that ZachS was secretly Kasparov himself, teaming up with Russian military-grade computers.

ZachS played creative and aggressive chess, ultimately winning the competition and revealing the mystery behind their identity. To everyone's surprise, the strongest AI machines and top chess grandmasters in the world had actually lost to two unknown amateur chess players from New Hampshire, aided by average household computers. What they had was a "really extensive database that [Steven Cramton] worked on for four or five years" and a "really good methodology for when to use the computer and when to use our human judgement, that elevated our advantage," Zackary Stephen says.[20]

This story demonstrates that the ability of a human mind to wrap itself around a problem with the proper data visualization tools still possesses a more holistic view than what is capable by machines. We feel this is especially true given all of the challenges around contextualizing Big Data. While human intuition would conclude that this space would be arbitraged out by the juggernauts of the industry, success remains much more of a function of skill rather than resources. Similar to the example of ZachS, the creativity of the human behind the machine and the experience of knowing when and how to use the data is almost as important as the data science.

a. Man versus Machine

We believe that alternative data must be approached with a blend of man and machine. Data science and analysis is inherently quantitative, and we look to automate a lot of the process. Given the fundamental nature of alternative data, we look to quantify mental models and the logic from fundamental research (i.e. develop a mathematical model for the "gut feeling"). This helps reference and iterate on successful work, keep personal biases at bay and use technology to provide scalability. The process is about 60% quantitative for the 1798 Q strategy today, but we continue to seek ways to make the process as automated as possible over time (80-90%). That said, there will always be a need for a human element to act as co-pilot. Data is ultimately a small sample of users for a multi-billion dollar company, speaking to one aspect of the top-line growth drivers. It tells humans where to focus, what questions to ask, and of whom, but personal experience is key to optimize the models for current market biases driving stock prices and unquantifiable risks.

The integration of man and machine is not an easy marriage. We constantly battle human biases based on past experiences while avoiding being led astray by blind trust of machines. Finding the balance that works for each manager depends on their fundamental and technical background. This governs what thresholds of data history they can be comfortable working with and how an abstract data series can be applied from a fundamental lens. We believe it requires someone who can wear both hats, speaking to vendors as frequently as management teams.

There is only so much a model can infer about the future from looking back at price or other data. It helps us understand the fundamental topline trends of a company systematically, but the rest requires discretionary judgement. How does revenue surprise translate into earnings surprise? How does earnings surprise translate into valuation multiple? How does that translate into share price movements? What impact does the company's event path have on stock price? What are the potential risks around liquidity or implied volatility? Even with the models being correct, it is possible to lose money if the process does not anticipate how major policy making, geopolitical risks, or simple behavior shifts can dictate stock moves around earnings announcements. A manager needs to appreciate the drawbacks and weaknesses of data to determine the accuracy and relevance of the models going forward.

For example, in 2017 Whole Foods Market went from being a growth organic foods stock to an M&A situation when Jana Partners got involved. While our data on their foot traffic and consumer spending continued to give the AI confidence in a earnings miss, the stock was no longer trading on earnings as much as it was trading on M&A speculation. Data can hint at recent revenues, but not what part of the earnings announcement the stock price will react to most. For that, human intuition is critical for position sizing decisions and risk management.

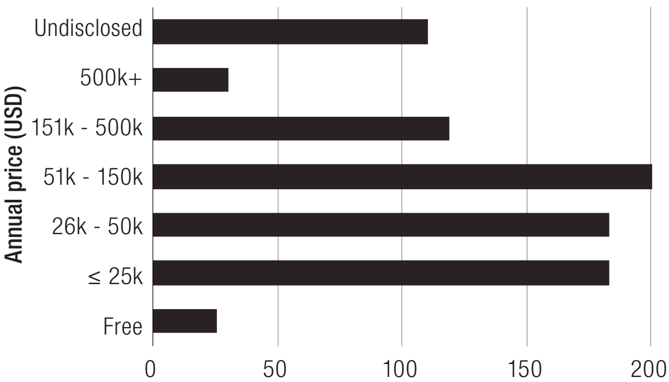
[20] BBC.com, "The cyborg chess players that cant be beaten" Chris Baraniuk, 4 December 2015.

b. The resource misconception

Given what we have seen around the bifurcation in the quantitative space (where the largest managers continue to grow), we are regularly asked how we can possibly keep up with their resources, size, scale, computing power and large list of PhDs on the payroll. Perhaps one of the largest misconceptions when it comes to alternative data is the belief that an investment manager's edge is related to their ability to access the most expensive, robust, or niche datasets and process them in-house.

There is no doubt that acquiring datasets and the talent necessary bears material costs, and that funds using alternative data will have higher AUM thresholds to break even relative to traditional equity long/short strategies. That said, if done effectively, this number can be much lower than investors anticipate. Growth of the ecosystem continues to tilt the scales and we argue that the investment edge in Big Data is more aptly linked to the investment manager's expertise and creative application of alternative data, than the size of their data science budget.

FIG. 14 COST DISTRIBUTION OF VARIOUS ALTERNATIVE DATASETS



Source: Neudata, Industry trends presentation, November 2018. Figures as of 12 October 2018 for 579 datasets.

Infrastructure:

- **Data science team –** In the early days, investment managers believed there was incremental value to be added by staffing full time data scientists in house to structure datasets. Today however, we can leverage the deep level of specialization within the external vendor ecosystem. These vendors often have dozens of data scientists dedicated to a single dataset and we find that the quality of tagging done externally tends to be better in the vast majority of cases. We believe the attitude of "if it is not built internally we are not going to use it" needs to be re-examined and the benefits of a large in house team do not justify the incremental expense and time of reinventing the wheel.
- **Compute capacity –** Legacy firms insist on having their own hardware and datacenters, but with the advent of cloud computing, we can scale and size our budget according to our needs. We believe shifting legacy models to cloud computing and open source software allows us to be more nimble and faster to market, while incurring lower up-front costs.

Data budget:

- **Buying power –** Once extracted, the data that vendors sell has no incremental cost to them, which makes for a negotiable price tag. Price can vary greatly based on willingness to accept reduced frequency, slight lag, or a smaller subset of data that may be sufficient. There is also buying power associated with experience in the space. Having met with hundreds of vendors over the years means that for any particular problem we are looking to solve for there are multiple solutions we can turn to, giving us more negotiating power.
- **Creativity –** There is no linear relationship between the cost and complexity of a dataset and the value that can be derived from it. Some of our most value additive datasets are actually some of the simplest and cheapest. As shown in Figure 14, the majority of datasets are actually at a much lower price point than many believe. With the proper evaluation tools and some creativity, managers can blend multiple smaller datasets across different types of data to paint a much more robust picture about a company and generate more alpha. We believe this approach lets us extract quality insights at much lower costs, while leaving out the most expensive datasets that do not add material value.

c. Maintaining a lead

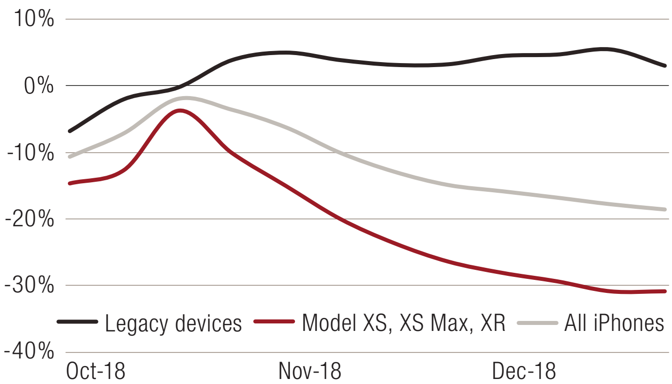
One question we get asked regularly is "how long can this last?" Surely if there is an informational arbitrage involved in Big Data it can't last long… right? We believe the degree of expertise and sophistication necessary to succeed in this space will actually limit this field to a smaller handful of participants than investors believe. While there will certainly be increased adoption of Big Data within the investment community and our informational edge in some of the more mainstream data sources may narrow somewhat, investment habits change slowly and this adoption will not be overnight. We continue to innovate how data shapes our understanding of business models, expanding our scope in terms of accuracy and time horizon. Furthermore, as data adoption spreads to new sectors and regions, we will also continue to grow our investment universe and identify new opportunities.

A head start in this space is critical and we believe the key to maintaining the lead is constant refinement of the systems and a focus on scalability. As exciting as developments in Big Data have been, we are still in the very early innings. Trending developments in technology like the Internet of Things (IoT) will continue to drive data complexity and bring rise to new forms and sources of data. The next five years will bring 4-5x growth in the amount of alternative data available to us, but it will most certainly also grow in complexity beyond the capability of today's systems. This means that managers with scalable research platforms, that can quickly normalize different schema and new datasets, will be able to continue to maintain an advantage faster than it is arbitraged away.

What are some examples of applying Big Data?

Apple Inc. ("AAPL") – Data exposes kinks in the armor
Data shows us where to spend our resources and where and when to ask questions. For AAPL, we may look at things like foot traffic in stores, credit card transactions, iOS store downloads, supply chain issues and iPhone activations. We know in AAPL's case, the vast majority of their valuation stems from a single product: the iPhone. We use multiple data sources to track iPhone demand in real-time. In a recent example, we were able to detect that appetite for the latest generation of iPhones (XS, XS Max and XR) was running meaningfully below market expectations and used this information to build a short position ahead of the surprising and rare negative pre-announcement to start 2019. While this pre-announcement forced the market to revise F1Q19 estimates down to more realistic levels, we believed the cycle of downward earnings revisions could continue. The level of granularity our data provides pointed to a few very interesting trends that go beyond the F1Q19 miss. We could see unit sales trends deteriorate materially as F1Q19 progressed, with November worse than October and December worse than November (Figure 15).
We also saw a shift away from the higher priced XS, XS Max and XR devices towards lower end/legacy devices, which carry far lower average selling prices and margins.

FIG. 15 CUMULATIVE IPHONE SALES YOY GROWTH DETERIORATED THROUGH F1Q19



Source: Proprietary LOIM data analysis using the 1798 Q "IDEA" Research Platform.

Apple and iPhone are a registered trademark of Apple Inc.

Facebook ("FB") – "Numbers don't lie" and other fairy tales

Conversely, data can also be misleading, but correctly identifying misconceptions can create opportunities from controversial situations. An example of this is FB, which suffered in the wake of the spring 2018 Cambridge Analytica scandal. After being in the headlines around the controversies of sharing user data, FB stock fell over 20% from the highs. Waves of negative press, potential regulatory backlash and the #quitfacebook movement on Twitter made it a battleground situation that was largely avoided by hedge funds into earnings. The popular impression, echoed in consumer surveys like the one recently conducted by Pew Research Center (Figure 16), appears to be that the negative publicity stemming from these disclosures will cause users to shun social media platforms, leading to lower profitability down the road.

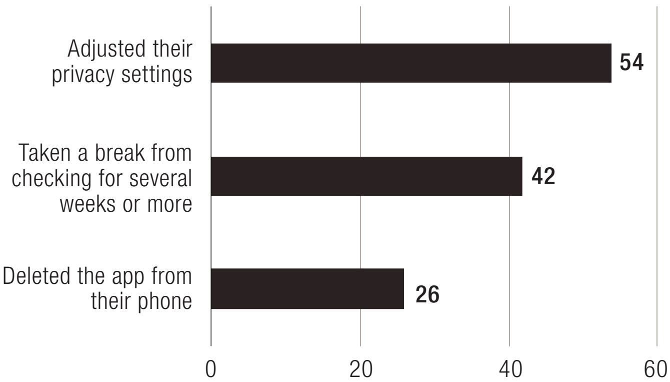
Beneath the surface of the outrage however, alternative data helped us confirm that what we want people to see and what we really do are often two very different things. As Seth Stephens-Davidowitz writes in his book *Everybody Lies*, "People can claim they're furious, they can decry something as distasteful, and yet they'll still click."[21] Looking deeper into FB's monthly active users with the help of data, we found that average active session times and FB application usage rates actually indicated no meaningful slowdown in utilization, or impact on ad revenues. Despite negative populous opinion, data helped highlight that the risk of FB missing their 1Q18 earnings was actually widely overstated. Data could not capture policy risk however, which is where the human element comes in with respect to additional research and risk considerations that go into position sizing.

Fading interest in home-meal kit trends[22]

A great example of the manner in which we combine insights from alternate data with deep fundamental knowledge of companies was our bearish bet on the home-meal kit delivery industry. Pre-packed meal subscription offerings from entrants like Blue Apron ("APRN") and HelloFresh ("HFG") and a host of other VC funded startups have seen sharp growth over the last few years, but profitability has always proven elusive. Many of the participants have argued profits would follow with scale, and as such have invested aggressively in marketing and promotions to grow their customer base. Using credit card data, we were able to pre-emptively pick up a sharp slowdown in sales across the industry during 2Q 2018; a sign of earlier than expected maturity as well as growing competition from grocers, who have piled into the market with their own offerings. More critically, we were able to dissect the credit card data at an individual user level to understand how rapidly customers cancel their meal kit subscriptions following the first month and discovered that customers are not staying long enough for the companies to recover the marketing costs of acquiring them. In other words, they are losing money on each customer they add, in which case scaling the product would actually exacerbate losses, not alleviate them. As seen in Figure 17, individual customer retention for home-meal kit companies is meaningfully worse than other consumer subscription models.

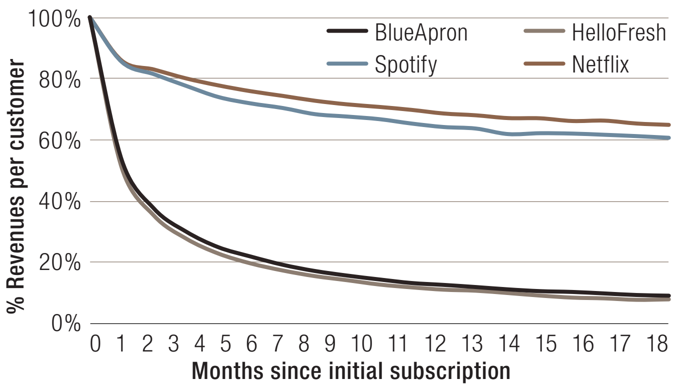
FIG. 16 FACEBOOK SURVEY SUGGESTS DECLINING USAGE

% of U.S. adults who use Facebook who say they have done the following in the last 12 months…



Source: PEW Research Center survey conducted 29 May – 11 June 2018.
Note: Those who did not answer or gave other responses are not shown.

FIG. 17 CUSTOMER RETENTION FOR HOME-MEAL KIT COMPANIES RELATIVE TO OTHER CONSUMER SUBSCRIPTION MODELS



Source: Proprietary LOIM data analysis using the 1798 Q "IDEA" Research Platform.

21 Everybody lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are, Seth Stephens-Davidowitz, 9 May 2017.

22 Proprietary LOIM data analysis using the 1798 Q "IDEA" Research Platform.

Final considerations

To reiterate, **Big Data is…**

1. **A game changer** that allows investment managers to gain insights and differentiated perspectives in order to enhance alpha and diversification for investors.
2. **Continuously improving,** growing in sheer volume, infrastructure and complexity. Investment processes must continue to evolve and innovate at this exponential pace.
3. **Tricky,** and to have any chance of success, investment managers must fully align their investment process and culture around the specificities and limitations of this diverse new data.
4. **Exclusive,** requiring a unique set of skills. The degree of expertise and commitment required for success will limit this field to a smaller handful of participants than investors think.

We are excited to see what this continued growth will bring for our space. Alternative data will continue to develop, offering more history, more vendors, better 3rd party off the shelf solutions and pricing efficiency. We look forward to continuing to adapt with it, making smarter more scalable algorithms and gaining new insights from evolving data and cultural trends. Big Data is a game of skill and scalability.

We hope this piece was insightful and we encourage readers to reach out to the 1798 Investor Relations team should they wish to learn more.

1798 Alternatives

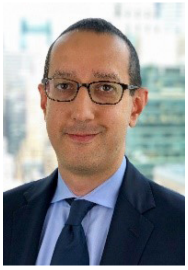
Lombard Odier Investment Managers (LOIM) is the asset management business of the Lombard Odier Group, with over 50 years of experience managing institutional assets and roughly USD 50 billion under management.

1798 Alternatives is our hedge fund and liquid alternatives group based out of New York (primary), London, and Geneva. Since launching our first hedge fund under the 1798 brand name in 2007 and having developed a robust hedge fund infrastructure, we decided to expand our offering in 2015 with two clear goals:

- Provide institutional investors with an alternative to large asset gatherers and crowded strategies.
- Provide pedigreed talent with expertise in innovative or capacity constrained strategies with the resources to succeed.

We believe that fostering pedigreed (20+ years experience) emerging managers within a collaborative investment culture backed by a 10+ year-old institutional infrastructure helps to elevate the insights and focus necessary to generate alpha in today's environment. Today, 1798 Alternatives oversees roughly USD 4 billion in client assets across 8 investment strategies.

Key contributor



Qaisar Hasan is the Portfolio Manager for the 1798 Q Strategy, which combines rigorous data science with fundamental analysis to construct a robust equity market-neutral portfolio. Before joining LOIM in May 2018, Qaisar was a Portfolio Manager for Point 72 Asset Management, where he launched the firm's data-driven long/short investment strategy. Qaisar is an industry leading expert in the development and application of Big Data analysis within alternative investment strategies. Prior to joining Point 72, Qaisar was a leading TMT analyst advising growth equity teams for Alliance Bernstein. During his 18 years of sell-side research experience, Qaisar has earned awards as the #1 coverage analyst in his sector from both Institutional Investor (2000)[23] and the Wall Street Journal (2008).[24]

Author

Gene Getman is a Client Portfolio Manager for the 1798 Alternatives group, focused on U.S investor relations for the Hedge Fund business. He is the Product Specialist for the 1798 Q Strategy and other innovative or capacity constrained alternative strategies based in New York. Before joining LOIM in April 2013, Gene was an Analyst on the Barclays Capital Solutions team, where he worked with Pensions, Endowments, Foundations and their consultants to make capital introductions to hedge funds. He began his career working for the NYSE Market Makers group within Lehman Brothers and Barclays Capital.

[23] Source: Institutional Investor, May 2000, "The 2000 all-Asia Research team," page 81.

[24] Source: The Wall Street Journal, May 2008 Telecommunications analyst rankings.

IMPORTANT INFORMATION

For professional investor use only.

This document is issued by Lombard Odier Asset Management (USA) Corp, a corporation organized under the laws of the State of Delaware, USA, having its principal place of business at 452 Fifth Avenue, New York NY 10018 USA.

Lombard Odier Investment Managers ("LOIM") is a trade name.

This document is provided for informational purposes only and does not constitute an offer or a recommendation to purchase or sell any security or service. It is not intended for distribution, publication, or use in any jurisdiction where such distribution, publication, or use would be unlawful. This document does not contain personalized recommendations or advice and is not intended to substitute any professional advice on investment in financial products. Before entering into any transaction, an investor should consider carefully the suitability of a transaction to his/her particular circumstances and, where necessary, obtain independent professional advice in respect of risks, as well as any legal, regulatory, credit, tax, and accounting consequences.

This document is the property of LOIM and is addressed to its recipients exclusively for their personal use. It may not be reproduced (in whole or in part), transmitted, modified, or used for any other purpose without the prior written permission of LOIM. The contents of this document are intended for persons who are sophisticated investment professionals and who are either authorised or regulated to operate in the financial markets or persons who have been vetted by LOIM as having the expertise, experience and knowledge of the investment matters set out in this document and in respect of whom LOIM has received an assurance that they are capable of making their own investment decisions and understanding the risks involved in making investments of the type included in this document or other persons that LOIM has expressly confirmed as being appropriate recipients of this document. If you are not a person falling within the above categories you are kindly asked to either return this document to LOIM or to destroy it and are expressly warned that you must not rely upon its contents or have regard to any of the matters set out in this document in relation to investment matters and must not transmit this document to any other person.

This document contains the opinions of LOIM, as at the date of issue.

The information and analysis contained herein are based on sources believed to be reliable. However, LOIM does not guarantee the timeliness, accuracy, or completeness of the information contained in this document, nor does it accept any liability for any loss or damage resulting from its use. All information and opinions as well as the prices indicated may change without notice.

Source of the figures: Unless otherwise stated, figures are prepared by LOIM. Although certain information has been obtained from public sources believed to be reliable, without independent verification, we cannot guarantee its accuracy or the completeness of all information available from public sources.

Views and opinions expressed are for informational purposes only and do not constitute a recommendation by LOIM to buy, sell or hold any security. Views and opinions are current as of the date of this presentation and may be subject to change. They should not be construed as investment advice.

No part of this material may be (i) copied, photocopied or duplicated in any form, by any means, or (ii) distributed to any person that is not an employee, officer, director, or authorised agent of the recipient, without Lombard Odier Asset Management (USA) Corp prior consent.

The information presented herein is intended to be a summary only and Lombard Odier makes no representation as to the accuracy of such information. This presentation may contain targeted returns and forward-looking statements. "Forward-looking statements," can be identified by the use of forward-looking terminology such as "may," "should," "expect," "anticipate," "project," "estimate," "intend," "continue" or "believe" or the negatives thereof, or variations thereon, or other comparable terminology. Investors are cautioned not to place undue reliance on such returns and statements, as actual returns and results could differ materially due to various risks and uncertainties. In considering any transactional history or prior performance information contained herein, investors should bear in mind that transactional history or prior performance is not necessarily indicative of future results, and there can be no assurance that the Fund will achieve comparable results. Such descriptions are provided for illustrative purposes only and there can be no assurance that Lombard Odier will be able to implement its investment strategies in a similar manner or pursue similar transactions or be able to avoid losses.

The categorizations, estimates, quantifications, sensitivities and other information provided in this presentation may not reflect the criteria employed by Lombard Odier to evaluate exposure, risks, or trading strategies. No representation is made that the categorizations, estimates, quantifications, sensitivities and any other information in this presentation are complete or adequate, or that they would be useful in successfully limiting exposure or risk, identifying and/or evaluating profitable or risky trading strategies, or constructing a profitable or limited-risk portfolio. The presentation of the exposures or risks identified in this presentation is not intended to be complete. There may be other exposures or risks, of which Lombard Odier may or may not be aware, that would affect the performance or risks of an investment portfolio. In addition, trading strategies may involve, among other techniques, leverage, short selling and various derivatives, each of which entails separate and distinct risks.

Any benchmarks/indices cited herein are provided for information purposes only. No benchmark/index is directly comparable to the investment objectives, strategy or universe of a fund. The performance of a benchmark shall not be indicative of past or future performance of any fund. It should not be assumed that the relevant fund will invest in any specific securities that comprise any index, nor should it be understood to mean that there is a correlation between such fund's returns and any index returns.

The case studies provided in this document are for illustrative purposes only and do not purport to be recommendation of an investment in, or a comprehensive statement of all of the factors or considerations which may be relevant to an investment in, the referenced securities. The case studies have been selected to illustrate the investment process undertaken by the Manager in respect of a certain type of investment, but may not be representative of the Fund's past or future portfolio of investments as a whole and it should be understood that the case studies of themselves will not be sufficient to give a clear and balanced view of the investment process undertaken by the Manager or of the composition of the investment portfolio of the Fund now or in the future.