Eagle Alpha

www.eaglealpha.com

Alternative Data
Use Cases

Edition 4


Sonraí Malartach
Alternative Daten
वैकल्पिक डेटा
另类数据
Dati Alternativi
Alternatív Adatok
Alternative Data
Datos Alternativos
Alternatywne źródła danych
Альтернативные данные
ალტერნატიული მონაცემები
Sonraí Athrach
εναλλακτικά δεδομένα
Données Alternatives
Dados Alternativos
بيانات
代替データ
Eagle Alpha

Table of Contents

List of Figures

Section 1: Executive Summary

Purpose Of This Report

The purpose of this report is to provide asset managers with examples of how alternative data can be used. This report is the fourth in a series of reports published by Eagle Alpha.

What's New ?

The primary differences between Edition 3 (published on December 5th 2017) and Edition 4 (published on April 11th 2018) are threefold:

1. There is a new section that identifies datasets that are in demand based on Eagle Alpha's proprietary analytics.
2. There are ten new case studies. See page 36 for a list.
3. We removed a section entitled "is there any alpha left in alternative data?" because this is no longer a common question asked by asset managers.

Alternative Data

Definition

Alternative data is non-traditional data that can be used in the investment process. There are 24 categories of alternative data in Eagle Alpha's taxonomy including consumer transaction data, geo-location data and sentiment data. Our sophisticated clients have already identified 1,500 relevant datasets and agree with our prediction that there will be 5,000 datasets by the end of 2020.

Benefits

These datasets are being integrated into the investment process primarily because they provide a greater volume of data and information compared to traditional datasets, unforeseen insight and give managers a competitive edge.

Drivers Of Adoption

Alternative data is increasingly being integrated into the investment process by asset managers. The drivers of adoption include:

- Competitive dynamics / edge e.g. 78% of US hedge funds are using alternative data.
- Growing evidence of alpha in alternative data e.g. 90% of managers get an ROI on their spend on alternative datasets.
- Providers of AUM expect to increasingly allocate to managers who are utilizing alternative data.
- Risk of being at a strategic disadvantage in the medium to long term.

Tipping Point

Alternative data is not new. Over fifty firms have been working with alternative data for several years. 2017 was a turning point for the alternative data space because the broader asset management industry began integrating it into the investment process. In our opinion the alternative data space will 'cross the chasm' (the tipping point) by the end of 2018 / Q1 2019.

Datasets In Demand: Proprietary Analytics

Eagle Alpha provides clients with data that highlights which data categories and specific datasets are in demand. The data is based on two sources:

1. Click-through data from clients that use our database.
2. 1-on-1 meetings organized between data vendors and asset managers.

Click-Through Data

Analysis of usage of our quantitative clients between Q1 2017 and Q1 2018 highlighted a few interesting trends. For example, the employment, consumer transaction, ESG and trade data categories received the most interest (normalized versus overall category size) in Q1 18.

1-on-1 Meetings Data

Analysis of 1-on-1 meetings stats also shows a few interesting trends. For example, web-crawled data/pricing data has seen a significant increase since Q4 2017. In addition, employment and sentiment data have been the most consistent quarter on quarter.

Use Cases

Asset managers regularly ask us for examples of how alternative data can be used. Section 4 outlines use cases for each of the 24 categories of alternative data, by asset class and by manager type.

Category Of Alternative Data

For example, web crawled data is frequently used to track website e-commerce activity (prices, product listings, promotions, reviews), public commentary, press releases, changes in corporate websites and government filings.

Asset Class

The majority of use cases apply to the equity asset class. However, an increasing number of credit funds are using alternative data. For example, we outline how the Scientific Active Equity team at Blackrock uses alternative data for CMBS.

Manager Type

Several quantitative and discretionary managers have publicly disclosed some of their use cases. For example, we Worldquant discusses how it uses satellite data, social media data, sentiment data and trade data.

Forty Case Studies

Asset managers constantly ask us for case studies regarding how alternative data can be used. The majority of this report (pages 36 to 104) is dedicated to forty case studies that are based on third party datasets as well as Eagle Alpha proprietary datasets. Case studies are provided for each of the 24 categories of alternative data, by asset class and by manager type. A number of the case studies have been backtested and/or verified by third parties such as JPMorgan and Citi.

Category Of Alternative Data

For example, we tested an employment dataset where results showed the “Jobs Active” variable producing the highest and most consistent returns (6-8% yearly). See page 39 for the case study.

Asset Class

Macro funds can leverage a South Korean real-time export dataset to accurately track overall China exports. See page 95 for the case study.

Manager Type

For example, discretionary managers used online retail data to identify improving sell-through trends for Fitbit in the first half of 2017. See page 66 for the case study.

Eagle Alpha

Since 2012 Eagle Alpha has provided a full service solution that enables asset managers to obtain alpha from alternative data.

There are four parts to our solution:

1. Data Sourcing keeps clients on top of all relevant raw and semi-processed alternative datasets worldwide.
2. Data Analytics enables clients to interpret alternative data using tools, indicators and analyst insights.
3. Bespoke Projects enables clients to commission curated work for specific research questions.
4. Data Forum enables clients to keep on top of the evolving alternative data space.

Section 2: Alternative Data

Section 2.1: Definition And Benefits

We define alternative data as non-traditional data that can be used in the investment process.

For a more specific definition we define alternative data by identifying all of the different categories of alternative data - currently Eagle Alpha's taxonomy has 24 categories (see Figure 1).

Today there are currently 707 datasets in our database that are spread across these 24 categories. It is interesting to note that some of our clients have already identified 1,500 datasets that are relevant to asset managers. We forecast there being 5,000 datasets by the end of 2020.

In our view the word "alternative" is currently in vogue but will fall away. The term the industry will use is simply "data", it does not matter if its traditional or non-traditional (alternative).

Figure 1: Eagle Alpha's 24 Categories of Alternative Data

Source: Eagle Alpha

Quinlan & Associates[1] published a useful summary of the benefits of integrating alternative data into an investment process (see Figure 2). They identify five benefits:

1. Greater volume of data and information.
2. Unforeseen insight.
3. Competitive edge.
4. Fiduciary duty.
5. Efficiency.

[1] Source: 'Alternative Alpha: Unlocking Hidden Value in the Everyday', September 2017.

Figure 2: Benefits of Integrating Alternative Data into An Investment Process

	Alternative Data	Traditional Data	Description
Greater Volume of Data and Information	• Frequently updated • Shorter history • Wide breadth	• Sparsely updated • Longer history • Narrow breadth	• Large amounts of data are being generated constantly, leading to more available information for better analysis • Frequent updates mean asset managers can carry out analysis sooner, enhancing portfolio construction
Unforeseen Insight	• Wide breadth • Profound implications to multiple assets, or even industries	• Narrow breadth • Only provides particular information, normally only directly relevant to few assets	• New data provides information on more than just financial performance, leading to better trend and performance predictions • Discovery of hidden relationships can help devise investment strategies
Competitive Edge	• Requires investments and capabilities	• Available to all • Analysis can be conducted relatively easily	• Talent and technologies are needed to gain value from alternative data, giving asset managers with the right resources higher alpha-generation
Fiduciary Duty	• Make use of available information and data	• Make use of financial data and few conversations only	• Asset managers are in charge of making optimal decisions for clients, and hence should incorporate as much data and analysis as possible into the investment process
Efficiency	• Rapid and efficient research and analysis • Wide coverage	• Highly manual research and analysis • Narrow coverage	• Alternative data provides insight to multiple assets, and can replace certain aspects of the current research process • Research analysts can spend longer time on better modelling or investment strategies

Enhance alpha-generation

Source: Quinlan & Associates

Section 2.2: Drivers of Adoption

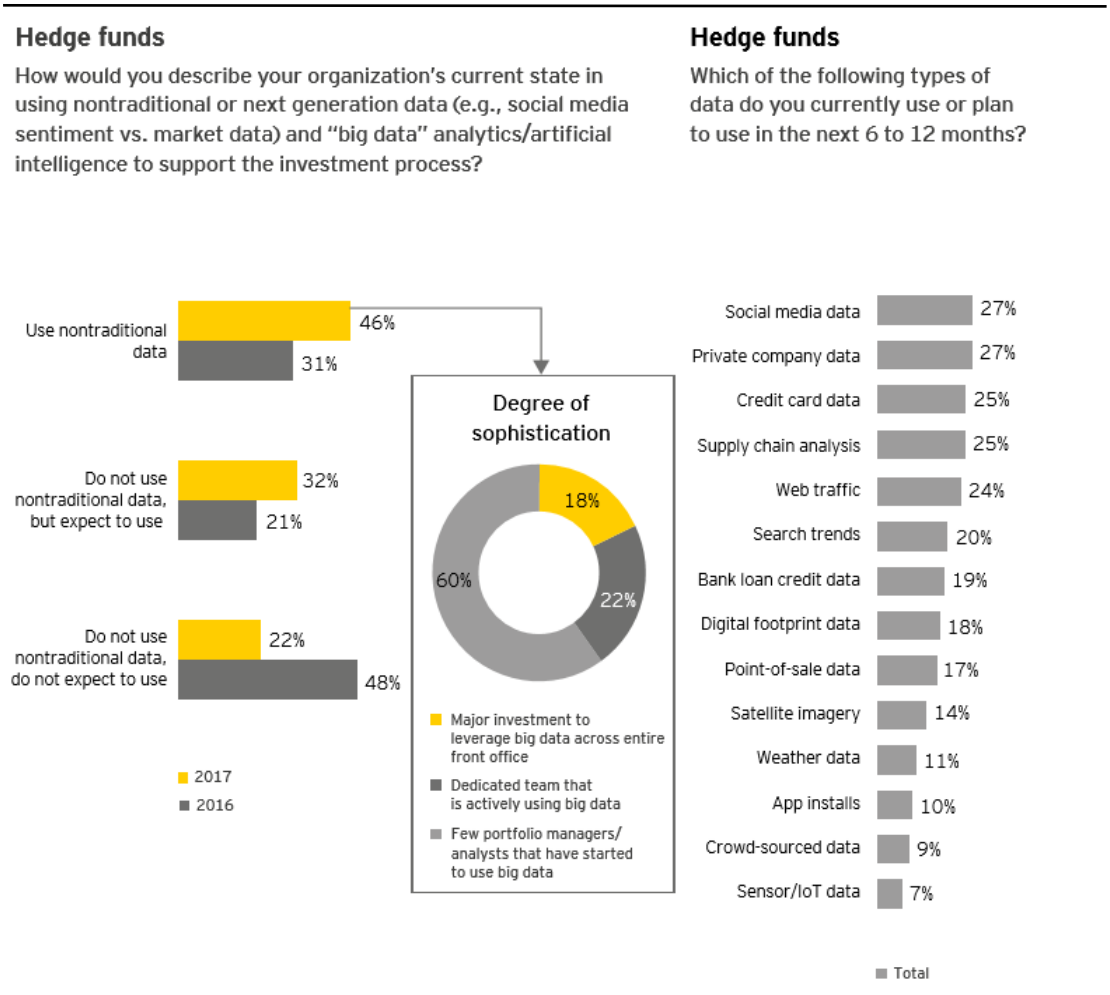
In Eagle Alpha's opinion there are four key factors that are driving the adoption of alternative data:

1. Competitive dynamics – 78% of hedge funds are using, or will start using, alternative data.
2. There is growing evidence of alpha in alternative data.
3. Investors expect to increasingly allocate to managers who are utilizing alternative data.
4. Risk of being at a strategic disadvantage.

1. Competitive dynamics – 78% of hedge funds are using, or will start using, alternative data.

Asset managers all have access to traditional data and information via traditional channels. Innovative asset managers are seeking an edge through non-traditional alternative data. In the US 78% of hedge funds are already using, or will start using, alternative data in the pursuit of an edge (Figure 3).

Figure 3: Excerpt from EY's Survey



Source: EY

The EY report stated:

- "A larger quantity of managers see effective use of data and analytics as a key competitive advantage for the future. Smaller managers moved first, but managers of all sizes and strategies are now experimenting with big data analytics and AI."

- "What is striking is how quickly the landscape has changed. Last year (2016) almost half of managers said they did not use, and did not expect to use non-traditional data in their investment process. However, in the current year (2017), 78% currently use or expect to use non-traditional data."

- "For 60% of managers "using non-traditional data and/or AI" means that they have a subset of their front-office teams experimenting with the tools. Fewer managers have fully dedicated teams or have made major infrastructure build-outs to support big data. As managers become more experienced and comfortable with these tools, we would expect to see continued evolution and sophistication around how managers incorporate big data into their investment strategy."

2. There is growing evidence of alpha in alternative data.

Greenwich Associates survey of asset managers[2], highlighted that 90% of asset managers that are using alternative data are seeing a return on their investment (Figure 4). Clearly measuring ROI on alternative data spend is not always straightforward. For example, some of our clients measure it as a multiple of spend whereas others measure it based on the number of interactions with a centralized 'Data Insights' team.

Other evidence that there is alpha in alternative data includes:

1. Manager PR: some asset managers are being vocal about their returns in the press (e.g. NN IP[3]).
2. Data vendors becoming funds: in some cases alternative datasets are so successful they became funds (e.g. Cargometrics).
3. Case studies: section 6 of this document provides 40 case studies that demonstrate alpha in alternative data.
4. More renewals: as an aggregator of alternative datasets we see an increasing number of datasets obtaining renewals of licenses.
5. Teams increasing: a headhunter that specialises in alternative data can show 50 asset managers that have steadily increased their alternative data headcount over the last few years.
6. Spend increasing: Opimas forecast spend on alternative datasets to reach $2bn by 2020 (see Figure 5).

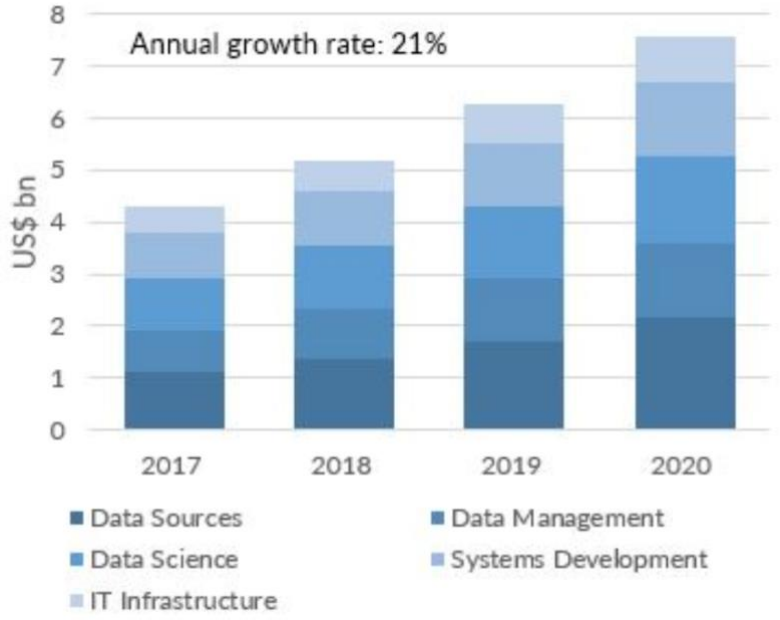
[2] Source: 'Putting Alternative Data to Use in Financial Markets', September 2017.

[3] Source: 'NN IP Gains From Sentiment Analysis', January 2017.

Figure 4: ROI From Alternative Data

Source: Greenwich Associates Survey in association with Arcadia Data

Figure 5: Spending on Alternative Datasets to reach US$ 2 billion by 2020



Source: Opimas

3. Investors expect to increasingly allocate to managers who are utilizing alternative data.

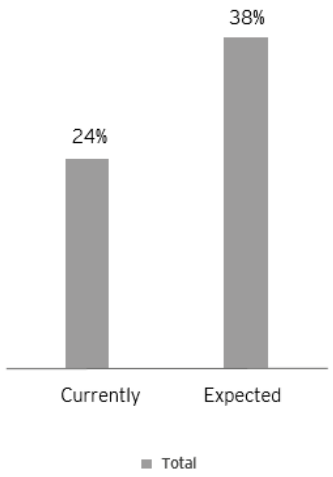
EY's survey of asset managers[4] highlighted that investors expect to increasingly allocate to managers who are utilizing alternative data – see Figure 6.

[4] Source: '2017 Global Hedge Fund and Investor Survey', November 2017.

Figure 6: Excerpt from EY's Survey

Investors

What proportion of the hedge funds in which you invest use nontraditional or next generation data and "big data" analytics/artificial intelligence to support their investment process? What do you expect that proportion to be in three years?



Source: EY

The EY report stated:

- "Investors are continuously searching for the next strategy, asset class or investment opportunity that they believe is unique and can outperform. Given the developments in FinTech and excitement surround the technological capabilities to rapidly analyze different datasets, it is not surprising that investors are expecting an increased percentage of their hedge fund managers to be using non-traditional data and new analytics in their investment processes."

- "Many investors view these advancements as an additional tool available to managers and those who are able to effectively harness the capabilities have a distinct advantage compared to those managers who are not deploying these capabilities within their investment process."

- "For years, these tools generally resided in the domain of quantitative managers. However, managers of all strategies have increasingly been innovative in developing methods to complement their investment strategy with these advancements."

4. Risk of being at a strategic disadvantage.

In September 2015 Blackrock's SAE[5] group stated "we believe that in order to generate sustained alpha, investors should embrace acquiring, analyzing and understanding the fast growing universe of data. Those that are unable to do so run the risk of falling behind in a rapidly changing investment landscape". Since then this view has gained acceptance amongst the wider asset management

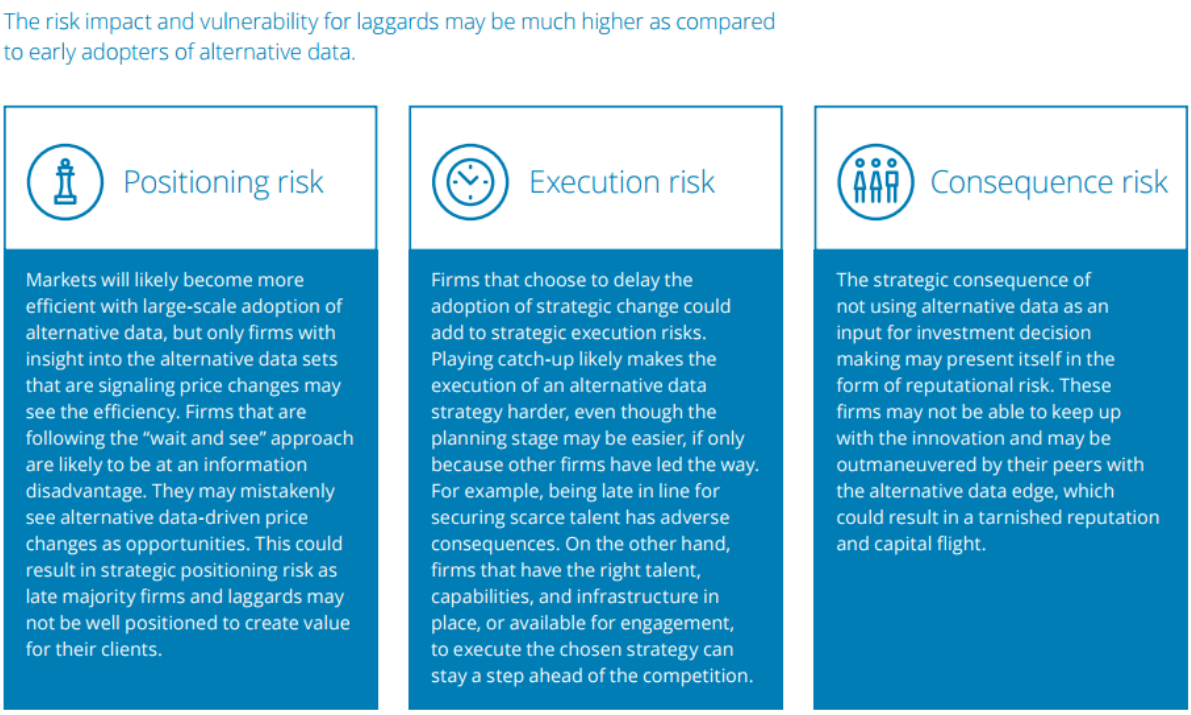
[5] Source: 'The Evolution of Active Investing. Finding Big Alpha in Big Data', July 2015.

vertical. For example, at JPMorgan's annual quantitative conference[6] in May 2017 "there was widespread agreement that big data and machine learning is transforming the investment landscape". In October 2017 Deloitte published[7] a report saying that firms that do not integrate alternative data into their investment process will be at risk of being at a strategic disadvantage.

The report stated:

- "Alternative data will likely transform active investment management over the next five years, from hedge fund management, to long-only mutual funds, and even private equity managers."
- "Those firms that do not update their investment processes within that time frame could face strategic risks, and might very well be outmanoeuvred by competitors that effectively incorporate alternative data into their securities valuation and trading signal processes."
- "The risk impact and vulnerability for laggards may be much higher as compared to early adopters of alterative data. There are three risks: 1) positioning risk; 2) execution risk; and 3) consequence risk." See Figure 7.

Figure 7: Risk Exposure Due to Late Adoption of Alternative Data



Source: Deloitte

[6] Source: JPMorgan research report entitled '2017 NY Macro Quant and Derivatives Conference', 25 May 2017.

[7] Source: 'Alternative Data For Investment Decisions', October 2017.

Section 2.3: The Tipping Point

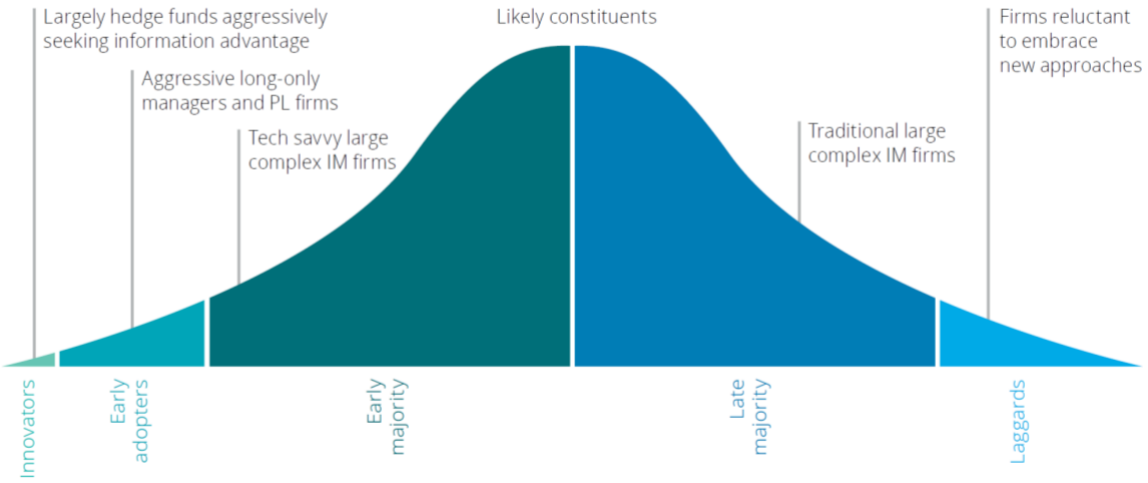
Alternative data is not new. 50+ innovative asset management firms have been working with alternative data for several years. These firms are primarily the larger quantitative and 'quantamental' firms. Jefferies, in its June 2017 paper entitled “Quantifying Intuition: Mapping the Data Science Landscape in the Hedge Fund Industry”, stated that 20% of hedge funds with over $1bn in AUM already have a person dedicated to alternative data or a person is spending 50% of their time on alternative data.

In our opinion, we are moving towards the end of the early adoption phase – see Figure 8. Based on data from our CRM database there are 250 firms that are seriously working with alternative data today. Our definition of “seriously” is where an asset manager has at least one individual dedicated to alternative data. The majority of these firms are in the US. However, other regions are starting to catch-up. For example, we are aware already 40 discretionary firms in APAC that have at least one individual dedicated to alternative data.

Adoption, and awareness, increased rapidly in 2017. We anticipate the alternative data space will ‘cross the chasm’ (the tipping point) by the end of 2018 / Q1 2019. This view is supported by Jefferies in their recent report entitled “Quantifying strategy: building data strategies for the net”. The report stated:

- “We are in the early stages of investment managers’ building out institutional data strategies”.
- “Even large organizations with dedicated headcount and resources consider themselves in the beginning of a multi-year secular shift”.
- “We believe that 2018 will be a period of continued transition and growth for data adoption among investment managers – and we expect that by 2019, the landscape for data adoption within organizations will be redefined”.

Figure 8: Alternative Data Adoption



Source: Deloitte

Section 3: Datasets in Demand: Proprietary Analytics

Section 3.1: Database Clickview Analytics

Eagle Alpha extracts insights from the aggregated click-through data of our global client base. User browsing behaviour can be used to identify data sourcing trends which reflect changes in buyside interest towards different alternative data categories and specific datasets. The product roadmap for our Data Sourcing solution includes the provision of this aggregated, anonymized data through on our online database. The insights and statistics will be presented for fundamental users, quantitative users and overall, to provide a holistic view of Data Sourcing trends.

Eagle Alpha's Data Sourcing team analyzed a subset of quantitative user's data between Q1 2017 and Q1 2018, identifying some interesting trends. The employment, consumer transaction, ESG and trade categories received the most interest (normalized versus overall category size) in Q1 2018. Open data, public sector data, social media data and pricing data received the least.

We observed strong year-over-year (YoY) interest (+4.24%) in employment data in Q1 2018, with sequential increases between Q3 2017 and Q1 2018. This is primarily driven by the entry of two strong vendors to the category which cover North American and U.K markets.

Despite strong interest in the consumer transaction category, we observed a decrease in click-through interest of 1.39% YoY in Q1 2018, following sequential declines between Q3 2017 and Q1 2018. We attribute this to a maturing of the consumer transaction data market, particularly in the United States which is well served by several high-quality providers.

ESG is a category which we have observed a material increase in interest from both our quantitative and discretionary clients. The category is appealing to quantitative funds due to its breath of coverage, and low correlation with traditional smart beta factors (please refer to case studies #4 and #5).

Trade data is also experiencing robust growth in interest, YoY +3.59% in Q1 2018. We attribute this to several new market entrants with coverage in North America and APAC, coupled to the acquisition of an established market participant by a large traditional data market incumbent.

Conversely, we have observed a significant decrease in click-through interest in geo-location datasets, down over 3% YoY in Q1 2018. We accredit this to a realization that geo-location data has still some way to go before maturing as a category, with rapid panel growth and a lack of history representing major challenges to working with this type of data.

Clients of Eagle Alpha are provided data showing which specific datasets are obtaining the most, and least, clickviews.

Figure 9: Insights From The Click-Stream Analytics

Row Labels	Q1 2017	Q2 2017	Q3 2017	Q4 2017	Q1 2018	Q1 YoY	Quarterly Avg	Q1 2018 Relative -/+
Employment	2.26%	3.01%	2.71%	3.11%	6.50%	4.24%	3.52%	5%
Consumer Transactions	7.34%	4.51%	11.31%	7.98%	5.95%	-1.39%	7.42%	3%
ESG	3.95%	3.01%	2.26%	3.49%	5.22%	1.26%	3.59%	3%
Trade	2.82%	3.01%	8.14%	4.82%	6.41%	3.59%	5.04%	3%
Reviews & Ratings	6.21%	2.26%	0.90%	1.90%	4.95%	-1.27%	3.24%	2%
Advertising	5.65%	1.50%	1.36%	1.52%	3.21%	-2.44%	2.65%	1%
App Usage & Web Traffic	2.26%	3.76%	2.26%	2.47%	4.58%	2.32%	3.07%	1%
Event Detection	3.95%	9.02%	1.81%	3.74%	4.03%	0.07%	4.51%	1%
Web Crawled Data	2.82%	0.75%	5.88%	3.17%	5.22%	2.39%	3.57%	1%
B2B Datasets	1.69%	4.51%	4.07%	1.77%	2.20%	0.50%	2.85%	0%
Store Locations	1.13%	3.01%	1.81%	0.63%	1.19%	0.06%	1.55%	0%
Expert Views	1.13%	0.75%	1.36%	1.65%	0.55%	-0.58%	1.09%	0%
Sentiment	4.52%	3.01%	8.14%	3.80%	4.30%	-0.22%	4.76%	0%
Online Search	1.13%	7.52%	5.88%	1.27%	1.28%	0.15%	3.42%	0%
Consumer Credit	0.56%	0.00%	1.36%	1.58%	1.10%	0.53%	0.92%	-1%
Geo-Location	7.34%	9.77%	2.71%	6.97%	4.21%	-3.13%	6.20%	-1%
Internet of Things (IoT)	0.00%	0.75%	0.90%	1.08%	0.73%	0.73%	0.69%	-1%
Satellite & Weather	2.26%	2.26%	3.17%	10.08%	4.76%	2.50%	4.50%	-1%
Business Insights	7.34%	11.28%	6.33%	10.33%	9.34%	2.00%	8.93%	-2%
Data Aggregators	6.78%	9.77%	3.17%	7.10%	4.30%	-2.48%	6.22%	-2%
Open Data	2.26%	0.00%	3.62%	3.93%	3.94%	1.68%	2.75%	-4%
Public Sector	1.13%	0.00%	3.62%	2.60%	1.74%	0.61%	1.82%	-4%
Social Media	7.34%	2.26%	9.95%	4.25%	5.13%	-2.22%	5.79%	-4%
Pricing	9.60%	8.27%	5.43%	5.89%	5.22%	-4.38%	6.88%	-4%

Source: Eagle Alpha

Section 3.2: 1-on-1 Meeting Analytics

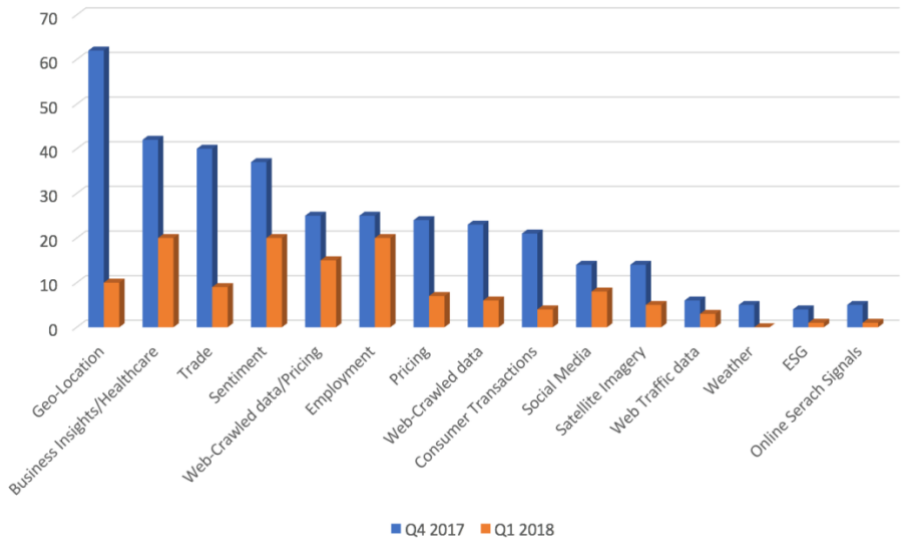
In addition to hosting events where alternative data vendors pitch to buyside firms we also continuously organise roadshows and 1-on-1 meetings throughout the year. Analytics on these meetings are made available to clients.

For example, web-crawled data and pricing data are two categies that saw a significant increase in meetings compared recently. In our opinion the key drivers behind this are:

1. Increased competition between data providers has led to very aggressive pricing.
2. Vendors have increased the quality of the data they are providing.
3. Increasingly asset managers are outsourcing web crawling projects in order to free up internal resources.

Clients of Eagle Alpha are provided data showing which specific datasets are obtaining the most, and least, 1-on-1 meetings.

Figure 10: 1-on-1 Meeting Analytics



Source: Eagle Alpha

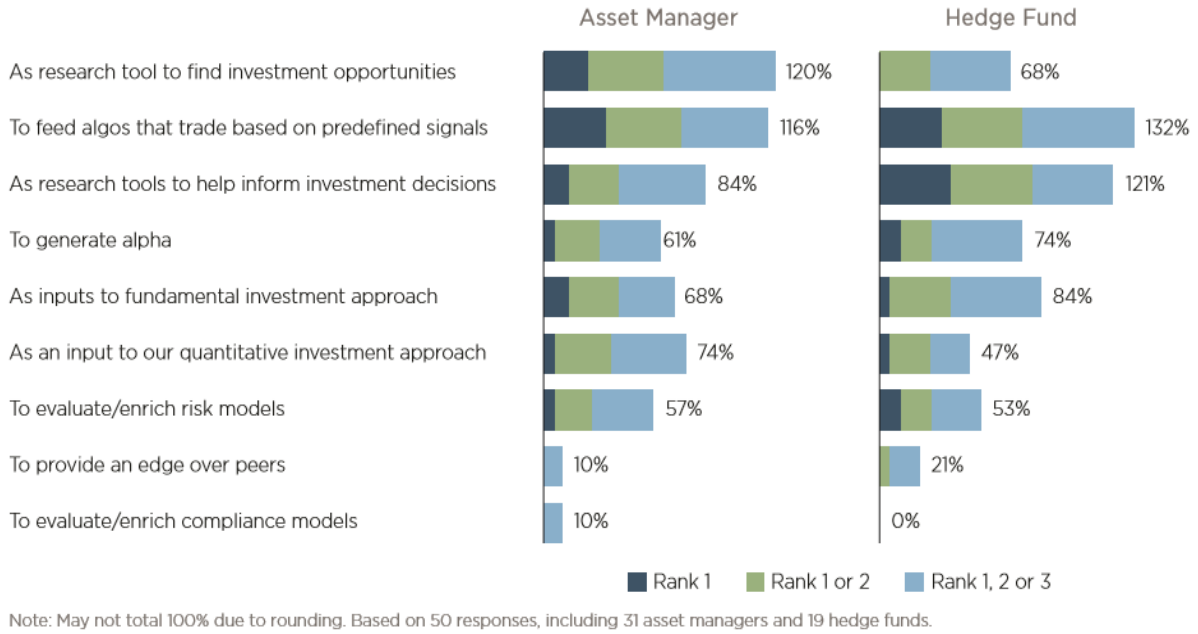
Section 4: Use Cases

Section 4.1: Overview

In this section we outline the different use cases of alternative data based on a Greenwich Associates survey in association with Arcadia Data. In addition, we summarize applications relevant to different categories of alternative data, asset classes (equity, macro and credit) and types of asset managers (quantitative funds, discretionary hedge funds and traditional fundamental asset managers).

Greenwich Associates survey of asset managers[8] in association with Arcadia Data shows that use cases trended towards specificity, not generalities (Figure 11) and towards details (specific anomalies, micro level analysis) – see Figure 12.

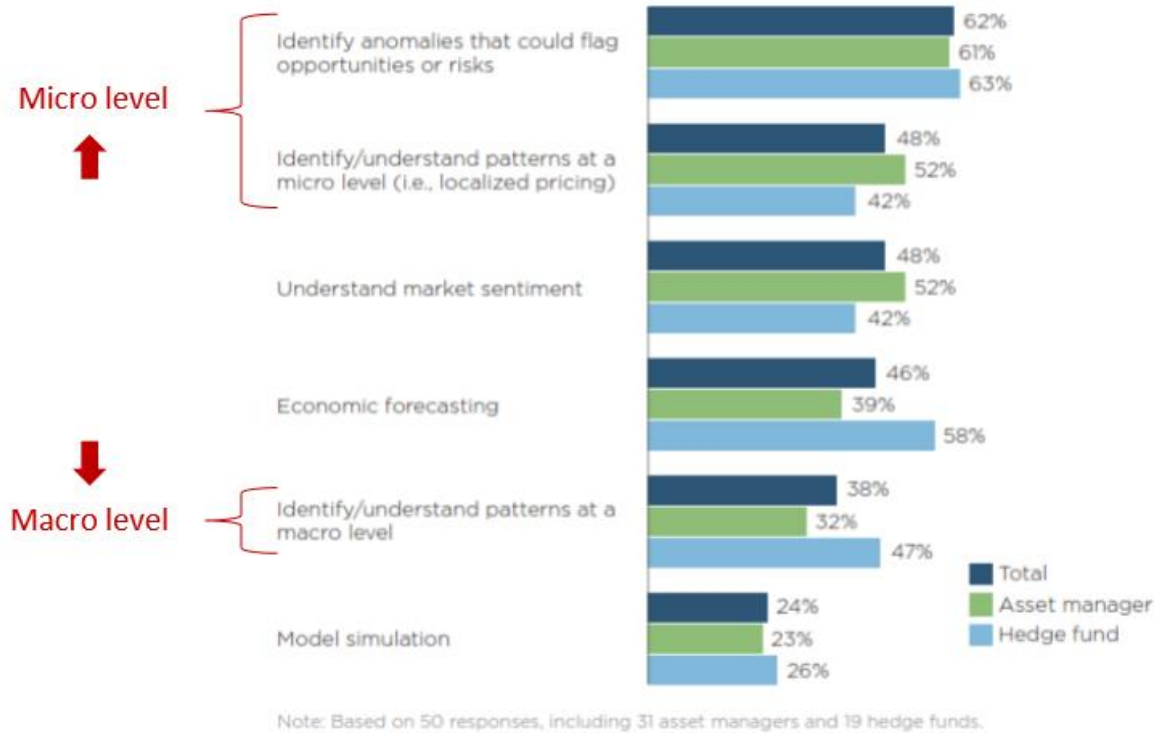
Figure 11: Using Alternative Data



Source: Greenwich Associates Survey in association with Arcadia Data

[8] Source: 'Putting Alternative Data to Use in Financial Markets', September 2017.

Figure 12: How Do You Use Alternative Data?



Source: Greenwich Associates Survey in association with Arcadia Data

Section 4.2: By Category of Alternative Data

Eagle Alpha created the first taxonomy for alternative data. This taxonomy has 24 categories – see Figure 13. The most common categories, based on our dialogue and analytics, include consumer transaction, geo-location, satellite and sentiment. As at 4th April 2018, there are 707 datasets in our database.

Access to detailed profiles on these datasets and insights into the most (and least) common datasets, based on our dialogue and analytics, are available to clients of Eagle Alpha's Data Sourcing solution.

Figure 13: Eagle Alpha's 24 Categories of Alternative Data

Source: Eagle Alpha

Below we highlight a number of examples of applications for each of the 24 categories of alternative data. These examples are based on our discussion with asset managers since 2012. For each category we state the number of relevant datasets in our database (as at 4th April 2018).

1. Advertising (21 datasets): data aggregators track corporate advertising spending on various platforms and by campaign. Advertising data exchanges have data on consumer interests over time based on their internet browsing habits. This data could be used to track category popularity e.g. luxury products and financial products such as mortgages, automobiles, cybersecurity etc. Very few investors have worked with this data and its predictive potential is largely unexplored.

2. App Usage & Web Traffic (39 datasets): web browsing traffic, both online and mobile, can be used to estimate company revenues, particularly if the web pages providing purchase confirmations are tracked. Mobile app usage can indicate the level of product adoption. Mobile banking, smart home devices, streaming media and food delivery apps are just a few of the potential use cases. Trends in mobile app reviews can help investors evaluate product success. Country specific data can provide insights into product adoption internationally. Investors can also track services embedded in apps such as payment providers and advertising services.

Please refer to the case study #25 on page 80.

3. B2B (21 datasets): a variety of data aggregators offer datasets about enterprise B2B commerce, including supply chain analytics. Some of these datasets are relevant for a range of industries, such as ones that monitor enterprise level internet browsing activity and Alibaba's B2B trade index. Other datasets provide niche information, such as databases of industrial materials and databases of oil contracts and drilling concessions.

4. Business Insights (114 datasets): certain datasets can provide unique insights into corporate activity. For example, by monitoring changes on corporate websites, new or discontinued services and mid-level management changes can be tracked. Other datasets provide insights into corporate credit quality. Datasets that detail acquisitions and startups at the industry and sub-industry level can yield insights into industry dynamics.

5. Consumer Credit (15 datasets): marketplace lending data is frequently updated daily, showing amount of loan issuance, loan pricing, borrower credit quality and defaults levels. Other data providers track consumer credit quality overall in specific countries. This data, which is typically more timely than other sources, can be used for determining momentum and inflection points in the consumer credit cycle.

6. Consumer Transactions (27 datasets): this data can come from a variety of sources and can provide merchant level transaction data (e.g. retailer, airline, service provider), product level purchase data (e.g. food, beverages, electronics) and macro level data. Some data sources, such as credit card transaction data, represent a large user base. Other data sources involve smaller panels, such as 2% of consumers, yet still provide reliable signals.

Consumer transaction data is frequently used to estimate quarterly revenue growth as the data is available before quarterly corporate earnings are released. However, consumer transaction data can also be used by long-term investors to gain insights into consumer purchasing behaviour. Example include rate of product adoption, trends in purchases of "premium" products, the effects of promotions and discounts, customer demographics and co-purchase behaviour. In addition, payment processing data, such as usage of PayPal and Square, is frequently identifiable in consumer transaction data.

In Q2 2018 Eagle Alpha will release its first semi-proprietary consumer transaction dataset. It is based on data from a partner that is a consumer transaction company.

Please refer to case studies #2 (page 41), #7 (page 49), #8 (page 50), #9 (page 52) and #21 (page 72).

7. Data Aggregators (63 datasets): technological innovation has allowed aggregators to collect data from disparate sources, and aggregate that data in a format that is helpful for asset managers. Aggregators may mine the deep web or carry out timely analysis of government filings and releases. Other aggregators operate exchanges, or platforms, where datasets may be purchased.

8. Employment (13 datasets): listings of job postings can be used to evaluate corporate strategy and direction, industry growth rates, and demand for specific skills. For example, is the demand for candidates with experience in Tableau, or Google AdWords, growing or plateauing?

Please refer to case studies #1 (page 39), #6 (page 48), #14 (page 59), #19 (page 68), #33 (page 92) and #36 (page 96).

9. ESG (20 datasets): alternative data sources can provide insights into the environmental, social and governance (ESG) standards at a company. ESG criteria may be monitored via a variety of data categories, including social media, satellite, open and public data. Additionally, sources which monitor business complaints, business reputation, employee compensation and hiring trends can also be useful resources.

Please refer to case studies #4 (page 44) and #5 (page 48).

10. Event Detection (36 datasets): alerts to breaking news from major news wires or social media sources allow traders to react before news is fully discounted in asset prices. Other events monitored include government filings and weather.

11. Expert Views (9 datasets): topic and sentiment trends among experts in any industry, or field of expertise, can differ substantially from the trends observed in the general population and in news feeds. The volume of information shared via niche blogs and forums make it difficult for investors to synthesize all of the commentary. Natural language processing (NLP) tools can aid in summarizing sentiment and topics.

12. Geo-location (48 datasets): location data derived from mobile devices can yield timely information on visitation trends. Common industry applications include retailers, restaurants, hotels, travel, transportation and REITs. In addition to observing the levels in foot traffic, this data can be used to identify the impact of promotions and weather events. Cross brand loyalty and regional idiosyncrasies may be identifiable. Geo-location data providers receive location data from mobile app owners, bluetooth beacons and sensors.

Please refer to case study #11 (page 56) and #16 (page 62).

13. Internet of Things (IoT) (14 datasets): while connected devices will eventually be able to provide data that allows asset managers to better understand consumer and business activity, the near-term opportunities lie in understanding which companies will benefit from IoT. Suppliers to the IoT industry and companies leveraging IoT in their course of business both stand to gain economic benefits. Data which tracks the digital footprint of IoT activity can provide indications of specific product adoption and overall market growth. As a nascent category, there are relatively few data providers, but IoT market development is likely to result in increased data availability.

14. Online Search (18 datasets): numerous academic studies have been published establishing that data regarding the volume of online searches can be used as an indicator of economic activity, as well as an indicator of consumer interest in a product or topic. Furthermore, these studies show that the best indicators are generally built with data from a basket of terms as opposed to a single term or a small number of terms. Complex data science

Eagle Alpha's Search Signals dataset consists of a series of company revenue indicators constructed using Google search volumes for a company's product offering. Observing crossing points of various moving averages of an index provides an insight into search momentum for a company's products, and has demonstrated a relationship with reported company sales metrics.

We have found across our software coverage that the best moving average to use for our signal is a 12, 6 and 3 month moving average. This makes intuitive sense given the business model and the sales cycle of software

techniques are used to determine the most indicative search terms and the most effective model for combining those terms into an indicator. Online search data has several years of history, is available in a timely fashion and is notably broad in its topic coverage.

Please refer to case studies #9 (page 52), #14 (page 59), #19 (page 68), #22 (page 74), #23 (page 76), #24 (page 78), #28 (page 84) and #32 (page 90).

15. Open Data (76 datasets): A tremendous amount of data is becoming available as open data. CKAN, Comprehensive Knowledge Archive Network, is a non-profit registry of open data. CKAN prepares data and provides access to data in ways that make that data more discoverable and usable. The CKAN data management platform is in use by numerous governments, organisations and communities around the world. Examples of open data that are relevant to investors include:

- Open Charge Map API that allows users to access data on locations of electric vehicle charging stations.
- The Wayback Machine provides a historical archive of internet pages which may be useful when backfilling data for a web crawling program.
- The GDELT Project provides a platform that continually records the world's news media from nearly every corner of every country in print, broadcast, and web formats, in over 100 languages, and provides a historical archive of news media content.

16. Pricing (96 datasets): aggregated pricing data of goods and services for both businesses and consumers is now more readily available than it has been in the past. This data can provide insights into corporate revenues and industry dynamics. Alternative measures of inflation have been developed using web crawled pricing data. This category also includes real estate sales, leases and rentals.

17. Public Sector (54 datasets): government agencies publish large datasets that can be used to gauge both social and economic activity as well as the impact of regulation on the market. For example, the Consumer Finance Protection Bureau (CFPB) provides access to its database of customer complaints relating to financial services companies. TED (Tenders Electronic Daily), a supplement to the Official Journal of the EU, publishes 460,000 calls for tenders per year, for about 420 billion euro of value.

18. Reviews & Ratings (27 datasets): product and service reviews posted online can be harvested and analysed for ratings trends and frequently mentioned topics. Numerous academic studies have shown that consumers place credence in online reviews and that favourable reviews generally lead to increased sales. At the same time, excessive reviews and complaints can be signs of poor management. App reviews can provide insights into consumer satisfaction with app services such as mobile banking. Other data providers track brand reputations by incorporating a variety of sources that gauge consumer and B2B opinions.

19. Satellite (56 datasets): the interpretation of satellite images into data or intelligence is useful to asset managers on many fronts e.g. general economic activity, agriculture, mining, construction and real estate, shipping, oil and gas, project monitoring and retail.

Please refer to case studies #12 (page 57), #13 (page 58) and #31 (page 89).

20. Sentiment (50 datasets): scoring of news feeds and social media posts by sentiment and novelty is a popular data source, especially for quantitative funds, due to its relatively longer history and columnar time series structure. Sentiment scoring may be applied to investor commentary, consumer attitudes toward products and brands, or mainstream news feeds. Sentiment data providers, in addition to mapping articles to entities such as government agencies and publicly traded companies, may provide additional scores relating to topic novelty, relevance, price impact estimate, and momentum. Data can be applied to factor models or used ad hoc in momentum and contrarian trading strategies.

Please refer to case studies #3 (page 43) and #37 (page 97).

21. Social Media (94 datasets): clients of Eagle Alpha typically leverage social media data for analysing consumer trends, reception of product launches, brand popularity, customer satisfaction, product sales promotions, and corporate/customer engagement. Brands with a growing number of unique individuals who engage with that brand on social media have shown to have favourable sales momentum, and vice versa.

Please refer to case studies #26 (page 81) and #28 (page 84).

22. Store Locations (14 datasets): tracking store locations can yield insights into corporate growth and strategy, particularly when store hours and promotions are also tracked.

23. Trade (36 datasets): macro firms leverage new alternative trade datasets for balance of payment estimates, insights into major commodity markets, indications of national competitive advantages and indications of consumer strength. Stock focused strategies use trade data to gauge sales of companies whose products can be linked to imports/exports of specific goods and to analyze supply chain activity. Trade data can also be used to gauge activity of transportation companies and publicly traded ports.

In 2018 a new supply chain shipping dataset will launch. The dataset represents transactions which are subject to tracking or taxation by the US Customs and Border Patrol (CBP). In general, this covers shipments entering the United States via a maritime vessel through a federal port, along with a small number of exports and "pass-through" transactions tracked by the CBP. Unstructured data from the CBP is transformed into a normalized view with clear parent-subsidiary entity linkages, allowing investors to to expose, integrate, and consume the data in meaningful ways.

Contact us for an introduction to this data vendor.

Please refer to case study #22 (page 74), #26 (page 81) and #27 (page 83).

24. Web Crawled (51 datasets): crawling projects are frequently used to track website e-commerce activity (prices, product listings, promotions, reviews), public commentary, press releases, changes in corporate websites and government filings. Data can be collected in-house or by companies that specialize in customized data collection. Datasets containing historical crawled data have been accumulated by web crawling companies and data aggregators.

Please refer to case studies #10 (page 54), #17 (page 63), #18 (page 66), #20 (page 70), #28 (page 84) and #39 (page 100).

Section 4.3: By Asset Class

In this section we outline examples of applications for the equity, macro and credit asset classes.

Equity

Equity investors are the biggest users of alternative data. Whilst the consumer and technology sectors are the most common there are alternative datasets available for every sector. Below we highlight a few examples of applications across 11 sectors and also disclose the number of datasets within each sector based on our current database (as at 4th April 2017).

1. Consumer Discretionary (147 datasets):

- For retailers, restaurants, and online travel companies, consumer transaction data can provide near real-time indications of consumer spending as well as longer-term insights into consumer trends.

- For product and service companies online search data has proven valuable at identifying inflection points in consumer interest that ultimately lead changes in revenue momentum.

- Data crawled from e-commerce sites can also provide an insights into product pricing dynamics as well as product traction as some sites disclose sales or sales rankings.

- Social media data and sentiment scoring can aid in understanding customers' perceptions of brands and products.

2. Consumer Staples (87 datasets):

- Consumer transaction data can provide valuable insight into performance of individual brands and retailers. ePOS data providers are particularly valuable for staples as they traditionally have strongest coverage for supermarkets and other consumer staples retailers.

- Studies have shown that consumers who engage with consumer staples brands on social media tend to be active buyers of those brands. Social media data can provide indications of brand momentum, brand messages that resonate with consumers, and corporate strategy for customer acquisition.

3. Energy (35 datasets):

- Trade data can provide insight into both demand and supply over the short and long-term. Not only does trade data capture current energy exports and imports, it also can be used to track upstream equipment imports. This data can be used for macro and micro economic applications.

- Satellite imagery can be used to track storage tank utilization across the globe. There are data providers that provide comprehensive information on the US oil and gas industry offshore, onshore and fracking. There are similar data providers for the Canadian market. Satellite imagery can also be used to monitor the facilitation of oil fracking sites across the US.

- Sensor data can be used to track the flow of oil and natural gas through pipelines.

- Credit risk data can be used to evaluate the real-time credit quality of oil and gas service providers, which can come under stress due to cyclical factors. Similarly, employment data can be used to evaluate cyclical pressures or upswings.

4. Financials (61 datasets):

- Growth in peer-to-peer lenders and other non-traditional lending channels has created a wealth of data to better understand the health of consumer credit and take the temperature of the lending market.

- App data can give insight into the adoption of financial services especially amongst millennials.

- Crawling data from consumer complaint databases can give an early indication of scandals involving financial institutions. In the U.S., regulators were alerted to the irregularities with Wells Fargo accounts from complaints to the Consumer Finance Protection Bureau database, which is open to investors use.

5. Healthcare (50 datasets):

- Trade data tracks real-time imports and exports of medical devices, allowing investors to monitor product demand.

- Aggregators source data from healthcare customers to provide data on medical device / hospital supply procurement.

- Other firms provide data on clinical trials and drug pricing which can be used in conjunction with social media commentary surrounding clinical trials.

6. Industrials (66 datasets):

- Trade data enables analysts to track industrial company shipments such as specialty glass or autoparts. It also allows analysts to track the impact of trade tariffs and duties.

- Geo-location data provides real-time information on passenger air traffic.

- GPS and trade data track the activity of ships and ports.

- Satellite data firms have been known to monitor large factories and key infrastructure projects which may add production efficiency or capacity.

- Many industrial conglomerates are focused on niche markets that can be tracked by expert views and surveys of business customers. Due to technological changes, surveys have become significantly less expensive to administer.

7. Information Technology (60 datasets):

- Hardware analysts can track the supply chain of major technology companies as well as the shipment of finished products by using trade data.

- Search interest for a particular software company's products and services can be monitored and compared to revenue growth.

- Traction of internet companies and online sales can be analysed by looking at mobile app data and email receipt data.

8. Materials (37 datasets):
- Trade data can be used to analyse commodity flows on a global basis. This can be used for macroeconomic use cases as well as company fundamentals.

- Satellite imagery can be used to facilitate supply projections for agricultural commodities and to track mining and shipments of raw materials.

- Metal prices can be predicted by using satellite imagery to capture inventory levels. Declining inventory levels being seen as a lead indicator of price increases and vice versa.

9. Real Estate (31 datasets):
- Using geo-location data, it is possible to track footfall to major shopping malls. This helps measure the health of mall sales in an environment where online commerce is taking market share.

- Eagle Alpha has partnered with a UK housing data provider that offers data on UK house sales and rentals that covers 80% of the market.

- Some data providers have information on inventory of commercial properties, building availability for lease or sale, existing tenant information and historical trends on demographics, occupancy and lease rates.

- Real estate pricing trends in emerging markets countries can aid analysts in understanding the breadth of economic growth as well as inflationary pressures.

10. Telecommunciation Services (20 datasets):
- Mobile app analytics providers can provide data on market share changes and quality of mobile telco services.

- Online gaming growth and video downloads can be tracked by numerous providers.

- Sentiment analysis across a wide range of online content can inform consumer attitude to particular telecom service providers and the services they offer.

11. Utilities (26 datasets):
- Sensor data can be used to track the flow of natural gas through pipelines as well as the amount of electricity traveling over transmission lines.

- Trade data contains information on solar panel imports and exports.

- Eagle Alpha's Web Queries product can track discussions about regulatory issues such as permitting and pricing.

- Consumer complaints to government agencies and social media commentary can be used to track ESG issues for utilities.

Macro

Wells Fargo[9], in a paper in April 2017, gave a useful high level summary of applications related to macroeconomic analysis: "Big data could help analysts solve many modern-day puzzles, such as productivity growth and its living standard relationship, micro-foundations of macroeconomic models, consumer/firm/investor behaviour and many more. In addition, big data would help include what is often the missing link of demographics in many economic/financial theories such as the consumption function. For instance, the current consumption function employed to analyze consumers' behavior estimates an average behavior that does not distinguish consumers' behavior by demographic or geographic region. Another potential utilization of big data would be to improve current methods to estimate the state of the overall economy as well as different sectors' / regions' performances. For the financial world, big data would increase opportunities for profits and help manage risk more efficiently by incorporating broader information in risk modeling".

Below we highlight examples of more specific applications across 8 macro categories and also disclose the number of datasets within each macro category (based on our database as at 4th April 2018).

1. Commodities (46 datasets): in 2016 Cargometrics (a shipping data vendor) was so successful that it turned itself into a hedge fund. In August 2017 Maersk invested in Cargometrics. There are several similar datasets that can be used by commodity investors to track supply and demand dynamics. Satellite data is also instrumental in forecasting agricultural commodity supply/demand as well as other raw materials such as iron ore.

2. Current Account (47 datasets): one of Eagle Alpha's data partners provides a Trade Nowcasting dataset that gives estimates of imports and exports for over 10 countries including China and Brazil. The timely and detailed breakdown of trade accounts allows analysts to better identify sustainable competitive advantages at the country level. Foreign exchange transaction data can aid in estimating capital flows.

Real time export data can be provided on a weekly basis at an aggregated level. For example, one can accurately track China export data by assessing South Korean export data which is captured on a weekly frequency.

3. Financial Instruments (45 datasets): Data sources that detect breaking news via social media are highly relevant to financial asset prices. Similarly, sentiment analysis of news and blogs has been found to correlate strongly with interest rates and foreign exchange rates. Sentiment analysis of

[9] Source: 'Big Data Applications in the Economics/Financial World Part I: Opportunities and Challenges', April 2017.

central bank communications is an example of a highly tailored application that may have a better track record than more mainstream natural language processing of central bank press releases.

4. Housing & Real Estate (66 datasets): the Bank of England uses a dataset of online residential housing listings as this dataset provides a more timely and region specific picture of real estate activity compared to traditional sources. It also provides a more real-time indication of residential real estate financing. Similar datasets, and ones that capture rental rates, are available for numerous countries.

5. Inflation (49 datasets): MIT's Billion Prices Project ("BPP") collects prices from hundreds of retailers to construct inflation indices. BPP suggests that its index is a good measure for predicting the U.S. inflation rate. Numerous data aggregators collect online pricing for industrial as well as consumer goods, providing valuable insights into sources of pricing pressures and price deflation.

6. Labor Market (40 datasets): one of the most popular datasets used by investors provides timely and granular data regarding the labor market, particularly the U.S. labor market. This dataset has 8 years of history, is mapped to thousands of tickers, is updated daily and can be analyzed at various levels such as state, industry and job category.

7. Personal/Household Sector (53 datasets): consumer transaction data is used to forecast private consumption. Datasets with aggregated credit and debit card transactions are available at the regional and sector level for the US and China. These datasets are typically updated weekly, thus are timelier than government monthly figures. These datasets also offer a check on official government statistics.

8. Surveys/Cyclical Indicators (58 datasets): several asset managers have leveraged online search data to build indicators for topics such as housing and employment. For example, the Scientific Activity Equity team at Blackrock[10] published a paper that gave an overview regarding how it uses internet search data to predict U.S. retail sales.

9. Interest Rates: sentiment and Natural Language Processing can be used to analyse Central Bank releases. A hawkish or dovish assessment can be a strong predictor of yield curve shifts in the subsequent period following the release.

For detailed macro case studies please refer to case studies #31-38 on pages 89-99.

Credit

We have seen credit investors use consumer defaults data to gauge consumer credit, geo-location data to track distressed situations of retailers and analysis of news and social media for municipal bonds (e.g. local government bankruptcies). Below we detail specific examples regarding muni-bonds and commercial mortgage backed securities.

[10] Source: 'The Evolution of Active Investing. Finding Big Alpha in Big Data', July 2015.

1. Municipal Bonds: in August 2017, IHS Market published a note[11] regarding how alternative data can be used for muni bond analysis. Below are the selected highlights:

- "Municipal bond investors are beginning to examine more "non-traditional" datasets for making investment decisions and surveillance given the potential wide gaps between reporting periods".
- "We believe that the inevitable rising rate environment will take away some of the cushion that has recently allowed municipalities to refinance into lower debt payments, which will increase the demand for new sources of data to more effectively price and tier municipal bond risk".
- "Using Puerto Rico's ports to gauge government income...the correlation is 54% and the R-squared is 0.2908 between the four-month lagged ship count and actual monthly tax collections from January 2014 to June 2017".
- "Using auto registration data to determine demographic shifts. A proxy for population migration is the changes in the U.S. auto vehicle registrations…the migration of new luxury vehicles between states is one potential gauge for the movements of higher income and net worth individuals among states".

2. Commercial Mortgage Backed Securities (CMBS): The Scientific Active Equity team at Blackrock outlined[12] how alternative data can be used for CMBS:

- "On the surface, the universe of CMBS appears to be a rather unruly dataset – but beneath the surface are rich veins of data at the individual loan or pool level. By combining machine-learning techniques and portfolio manager expertise a deeper understanding of the underlying data can be gained".
- "In order to extract value from data on tens of thousands of CMBS loans, it is necessary to write programmes that can read important information that is generally contained in an annex to the prospectus supplement for each security. This is a complicated process with a high potential for error that requires strict quality control standards. In addition to the top line data that a well-designed algorithm can extract, CMBS can also contain subordinate financial details that are buried within sub paragraphs or footnotes of loan documentation. These can significantly alter the top line data".
- "The details of this kind of subordinate financial need to be assessed manually by experts with product specific knowledge in order to gain a more complete understanding of the risk profile of individual loans. If managers are able to combine the necessary human market expertise with a powerful computing platform, it is possible to build a model that can help to better predict loan default".

For a detailed case study regarding credit please refer to case studies #39 (page 100) and #40 (page 102).

[11] Source: 'Boats, Quotes, and Automobiles: Alternative Data for Municipal Bond Investors', August 2017.

[12] Source: 'The Evolution of Active Investing. Finding Big Alpha in Big Data', July 2015.

Section 4.4: Applications by Type of Asset Management Firm

In this section we outline examples of applications for quantitative funds, discretionary hedge funds and traditional fundamental asset managers.

Quantitative Funds

Many of the early adopters of alternative data were the larger quantitative funds. The use cases are varied - at JPMorgan's 2017 quantitative conference 237 investors were asked "How do you plan to use Big Data and Machine Learning in investing?"

- 30% said to enhance existing risk premia / quant strategies.
- 20% said to build new risk premia / quant strategies.
- 25% said to enhance portfolio construction / risk management.
- 17% said to make discretionary calls / trades.
- 8% said do not plan to use.

In a July 2017 paper entitled "Discover the Hidden World of Alternative Data" WorldQuant stated: "for investors, especially those who play in the world of quantitative finance, increasing access to alternative data from emerging sources like IoT could give rise to a variety of new investment ideas and trading strategies". Below we outline specific applications, based on public information, of WorldQuant, Blackrock SAE and Acadian.

WorldQuant[13]:

- Satellites "can keep track of the number of factories under construction in rural China, a possible indicator of the strength of that country's industrial production".
- "Social media has been a particularly rich source for analysis by investors who are trying to understand market sentiment and predict business performance".
- Sentiment "analysis is useful in determining new-product perception and brand reputation, assisting investors in forming predictions about the growth of a company or industry".
- "Real-time access to data on ship movements can prove useful in building a comprehensive picture of global shipping and an accurate understanding of competitor and market trends".
- "Acquiring diverse shipping data can be used to better understand the costs and health of companies' supply chains".

[13] Source: 'Discovering the Hidden World of Alternative Data', July 2017.

Blackrock SAE[14]:

- "One area of research that has proved rewarding is measuring the impact of employee sentiment on company profitability…Today by using big data analysis and crowdsourcing research, we can not only measure employee sentiment on a much wider scale, but also in a more timely and accurate way. To do this we automated 'web-scraping' capabilities to look into job sites where employees provide feedback on thousands of employers around the world, as well as other sources of employee sentiment including social media, blogs and chat rooms. When all of the relevant data from these sources has been collected an overall sentiment score for a company is calculated and this score can be regularly updated and changes monitored".

- "Rather than focusing on a 70-year-old monthly survey of 500 consumers (Michigan Consumer Sentiment Index), we think it is possible to achieve more accurate results with a more modern toolkit and a much larger sample size…Most consumers now precede big purchases with some research on the internet. By tracking search activity, around the world, or big-ticket items, we can gain a more complete picture of consumers' purchasing intentions, which also includes granular data on the types of items that are likely to be purchased in different geographic regions. The results of this type of analysis appear promising".

Acadian[15]:

- In March 2017 Acadian announced that it was the first investment firm to use Bing Predicts, a machine-learning project that mines internet search and social-media data, for factors to try to forecast events like a corporation's quarterly results.

- "With Bing search history, we know what consumers are searching for, and how that will relate to the future earnings of companies," Ryan Stever, Director of quantitative global macro research. "Anything that can speak to the future earnings growth of a company is going to be valuable to us".

Discretionary Hedge Funds

According to the Barclays[16] survey regarding alternative data, 24% of discretionary hedge funds use alternative data with the most popular categories being: 1) consumer credit card; 2) internet / social media; 3) weather; and 4) satellite. See Figure 14.

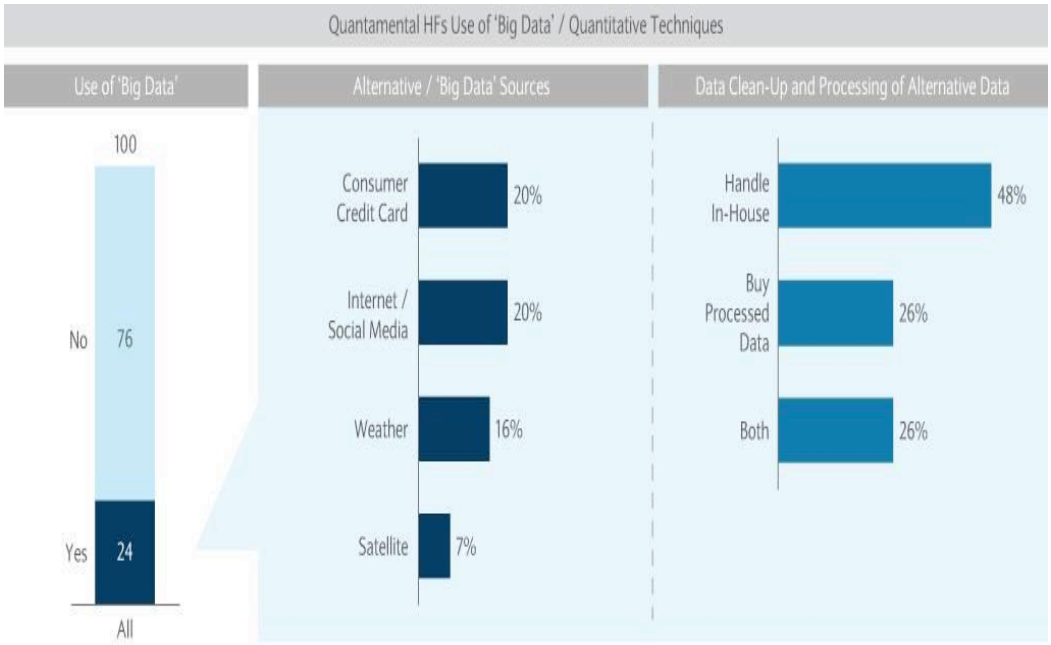
Increasingly, we observe discretionary hedge funds using alternative data for stock due diligence and for testing negative or "short" based investment theses. For example, is the management's representation of the business accurate or if sentiment around the business brand and products is shifting. In some instances this has been applied in regions where corporate disclosures and corporate governance is weak. This is applicable to both public and private markets.

[14] Source: 'The Evolution of Active Investing. Finding Big Alpha in Big Data', July 2015.

[15] Source: 'Acadian to use Microsoft's big data technology to help make bets', March 2017.

[16] Source: 'Rise of the machines', June 2017.

Figure 14: Results of the Barclays Survey Regarding Alternative Data



Source: Barclays

Below we give three examples of hedge funds that have publicly disclosed that they use alternative data:

- **Third Point**[17]: in its 2016 year-end investor letter stated: "we have added data science to our toolkit for identifying interesting, uncorrelated opportunities".
- **Point 72**[18]: at a CB Insights conference in June 2016 the Chief Market Intelligence officer of Point72, Matthew Granade, stated that alternative data is useful for generating alpha. He said: "it is a real change from how investing used to work…if you want to understand what is going on with McDonald's, you are going to have to look at credit card transactions data, you are going to look at geo-location data, at app downloads and a handful of other things. And suddenly you are going to have a very robust picture of how McDonald's is doing and you are not going to have to talk to McDonald's about that".
- **Citadel**[19]: the firm is building a centralised data organisation. Ken Griffin, founder and CEO of Citadel, stated "our ability to leverage big data effectively in our investment process is critical to our success as a firm".

[17] Source: 'Daniel Loeb's Third Point Hedge Fund 4th Quarter Commentary', February 2017.

[18] Source: 'The New Alpha: How Alternative Data is Going to Change Institutional Investing', June 2016.

[19] Source: 'Uber's Laszlo Korsos Joins Citadel as Chief Data Officer', July 2017.

Traditional Fundamental Asset Managers

A Greenwich Associates survey[20] regarding alternative data included responses from 46 traditional fundamental asset managers. When asked what were the benefit metrics of alternative data sources?

- 57% said changes to business risk predicted in the coming months.
- 46% said predictive indicators of future viability of a business based on its payment experience.
- 41% said risk of disruption of a business's supply chain.
- 41% said benchmark or index measuring the health of the small business economy at the national/regional/industry level.

Based on Eagle Alpha's dialogue with traditional fundamental asset managers, the five most common applications of alternative data are (in no particular order):

1. Identification of consumer trends and preferences: new products (Samsung S8), geographic expansion (e.g. Monster Beverages, Netflix), brand strength, and customer demographics.
2. Assess corporate quality from employee and customer reviews, social media commentary and government complaints. We have seen an increase in demand for ESG datasets.
3. Monitor industry competition: pricing, promotions and capital investments.
4. Evaluate corporate execution via website changes, store growth, employment data, and trade data.
5. Gauge pace of secular industry trends e.g. the electronification of automobiles and real-time payments.

Below we give three examples of traditional fundamental asset managers that have publicly disclosed that they use alternative data:

- **Schroders**[21]: in its 2015 annual report, Schroders stated that "analysis of 'big data' could become a key differentiator…this year we set up a Data Insights team, representing a significant new initiative for the Group. The team is focused on developments in data analytics for investment and research, to enhance and complement the existing skills of our fund managers and analysts". The report went on to say "the quantity of information available for investment research purposes is increasing at such a rate that traditional industry practices and skillsets are unable to absorb and process it. Global trends in digitalisation, social media, open data and technology are all creating vast streams of alternative data that are often highly unstructured and extremely obscure. However, they contain valuable and often unique insights".
- **State Street**: in an August 2016 Institutional Investor article[22] entitled "Unexpected risk meets unexpected data," State Street's Chairman and CEO highlighted a few ways investors can use different sources of information to enhance portfolio transparency and identify risk exposure ahead

[20] Source: 'Alternative Data for Alpha', January 2017.

[21] Source: '2015 Markets in Review', March 2016.

[22] Source: 'Unexpected Risk Meets Unexpected Data', August 2016.

of potential black swan events e.g. “Online retail. When consumers order products, they may be helping investors better track inflation trends to help recalibrate investment strategies before, and after, an event. PriceStats, an inflation series built by State Street Global Markets on online data, uses technology to monitor price fluctuations on roughly 5 million items and tends to identify price shocks faster than similar measures of offline prices, helping investors quickly understand potential shifts in inflation in more than 70 countries”.

- **NN Investment Partners**[23]: In an interview in May 2017 a senior portfolio manager (Mark Robertson) told Fund Selector Asia that “opinions and emotions expressed in online content, from news articles, through blogs, forum posts, social media such as Facebook or LinkedIn, to tweets, can provide a sense of market sentiment that can reinforce or even anticipate fundamental indicators, thereby helping make investment decisions”. Mark went on to say that “the big data we’ve incorporated is very good at picking out turning points and extremes”. The article included a case study:
    - In early 2017, the team’s scorecards were generally positive on commodity markets, despite some signals of credit tightening in China.
    - As part of its sentiment analysis, the team monitored political risk and emotional sentiment indices around commodities.
    - “Around mid-April we saw a really sharp deterioration in both of these signals in our scorecards, at the same time the overall scorecard was still giving us a positive view”.
        - The sentiment shift reinforced the misgivings the team already had, based on their fundamental analysis, and a decision was made to reduce commodity exposure.
        - The Bloomberg Commodity Index subsequently fell from 86.3 on 13 April to 82.0 on 9 May 2017.

[23] Source: ‘NN IP Gains From Sentiment Analysis’, May 2017.

Section 5: Forty Case Studies

In this section we present forty case studies that are based on Eagle Alpha proprietary datasets and tools as well as third party datasets. Our goal is to show case studies relevant to a variety of asset classes and investment approaches and which use a variety of alternative data types.

The case studies are in the following order: equity, macro and credit. Within each asset class the case studies are in the following order: quantitative use cases, short-term discretionary use cases and long-term discretionary use cases.

Edition 4 includes ten new case studies (#5, 6, 14, 15, 16, 29, 30, 36, 37, 38).

12	Equity (DKS)	Discretionary (short-term)	Satellite Imagery	Satellite data analysis showed falling traffic at Dick's Sporting Goods since the start of 2017. On August 15th 2017, Dick's Sporting Goods reported worse-than-expected results.	57
13	Equity (CMG)	Discretionary (short-term)	Satellite Imagery	Satellite data analysis showed falling traffic at Chipotle Q4 2014. This analysis was ahead of the street as sell side analysts only began revising Chipotle down 3 quarters later.	58
14	Equity (CMG)	Discretionary (short-term)	Employment, Online Search	Negative growth in active job listings and Google search data were used to correctly call worsening momentum for Chipotle.	59
15	Equity (AR)	Discretionary (short-term)	Sensor	The sensor data provider observed production increases at Antero Resouces and correctly anticipated management would raise company guidance.	61
16	Equity (RSH)	Discretionary (short-term)	Geo-Location	Geo-location data was used to anticipate disappointing sales at RadioShack stores.	62
17	Equity (GPRO)	Discretionary (shorter-term)	Pricing	Online pricing data pointed to negative fundamentals for GoPro, which were reflected in subsequent results and stock price.	63
18	Equity (FIT)	Discretionary (longer-term)	Pricing	Online pricing data showed improving sell-through trends for Fitbit in the first half of 2017. On August 2nd 2017, Fitbit reported better than expected results with adjusted revenue of $353.3m vs consensus estimate of $339.2m.	66
19	Equity (HUBS)	Discretionary (long-term)	Employment Online Search	Accelerating growth in active job listings and Google search data were used to correctly call improving momentum for HubSpot.	68
20	Equity (EXPE)	Discretionary (long-term)	Pricing	A web data provider accurately predicted that EXPE would miss 3Q17 room-night growth expectations, based on decelerating trends in reservation growth it started flagging in August.	70
21	Equity (SQ)	Discretionary (long-term)	Consumer Transaction	Analysis of the email receipt data for Square indicated that the growth in number of sellers has been in decline since the first quarter of 2010.	72
22	Equity (FINL)	Discretionary (long-term)	Online Search	The search indicator correctly predicted weakness in Finish Line sales ahead of earnings in December 2016.	74
23	Equity (BRBY)	Discretionary (long-term)	Online Search	Citi concluded that the short-term 1-month YoY observation crossing over the 3-month moving average YoY indicates major inflection points of same store sales growth for Burberry.	76
24	Equity (FINL, FL, NKE, UAA, ADS)	Discretionary (long-term)	Online Search	Online search data provided early indicator of weakness in sportswear sector. The industry analysis supported our case for fundamental weakness for FINL and FL and pointed to longer term fundamental issues for both companies.	78
25	Equity (TWX)	Discretionary (long-term)	Mobile App	App data showed an early indicator of a positive inflection in revenue growth for HBO, one of Time Warner's largest divisions.	80
26	Equity (ATVI)	Discretionary	Social Media	Using social media data, we correctly highlighted that the Overwatch game was well positioned to set a new sales	81

		(long-term)		record for Activision Blizzard.	
27	Equity (042670 KS, 013570 KS, 036890 KS)	Discretionary (long-term)	Trade	The South Korea real-time export data accurately tracked revenue of construction machinery companies throughout the Q1'12 – Q2'17 period.	83
28	Equity (LULU)	Discretionary (long-term)	Online Search, Social Media, Pricing	Eagle Alpha's analysis proved correct i.e. Lululemon reported sales growth of 13% YoY in Q3 2016 which was in line with our expectations.	84
29	Equity	Discretionary (long-term)	Online Reviews	Companies that get better reviews from employees post better share price performance.	86
30	Equity	Discretionary (long-term)	Online Reviews	The assessment of service and product quality of auto lenders has proven to be indicative of deeper operational issues which are then reflected in stock prices.	87
31	Macro	Discretionary (short-term)	Satellite Imagery	A relationship was identified between copper prices and estimates of copper inventories using satellite imagery data.	89
32	Macro	Discretionary (long-term)	Online Search	EA's US Unemployment index has a 5-year correlation of 0.9 with the US Unemployment Rate, with an out-of-sample prediction improvement of 14% compared with a baseline ARIMA model.	90
33	Macro	Discretionary (long-term)	Employment	Employment data enables more granular analysis of the labor market by sector.	92
34	Macro	Discretionary (long-term)	Trade Nowcasting	Statistical backtesting has proven the indicators are frequently a better predictor than street-mean estimates, while also having the advantage of a substantial lead time.	93
35	Macro	Discretionary (long-term)	Trade	The South Korea real-time export data accurately tracked overall China exports.	95
36	Macro	Discretionary (long-term)	Employment	Using payroll processor microdata improves forecast accuracy of an aggregate labor market activity measure.	96
37	Macro	Discretionary (long-term)	Sentiment	Quantitative metrics of emotional content in market narratives may complement other indicators and analysis in helping to gauge systemic risk.	97
38	Macro	Discretionary (long-term)	Trade Nowcasting	Trade data was used to track the surprising macro-economic turnaround for Brazil in 2016.	98
39	Credit	Discretionary (long-term)	Pricing	Online property listings data was used to research buy-to-rent investors which led to the subsequent clamp down on mortgage lending.	100
40	Credit	Discretionary (long-term)	Credit Risk	Testing showed that the probability of bankruptcy within 12 months ranges from 10% to 50% when a company gets a stress score of "1".	102

1. Equity > Quantitative > Employment Data

Key Takeaway

Testing shows that there is alpha in the data, with the "Jobs Active" variable producing the highest and most consistent returns (6-8% yearly).

Dataset

Completely unique in the industry, the job listing dataset only indexes jobs directly from employer websites. Updated daily with over 4 million jobs from more than 30,000 employers, the platform eliminates duplicate and expired job listings, as well as job pollution.

Backtesting/Significance

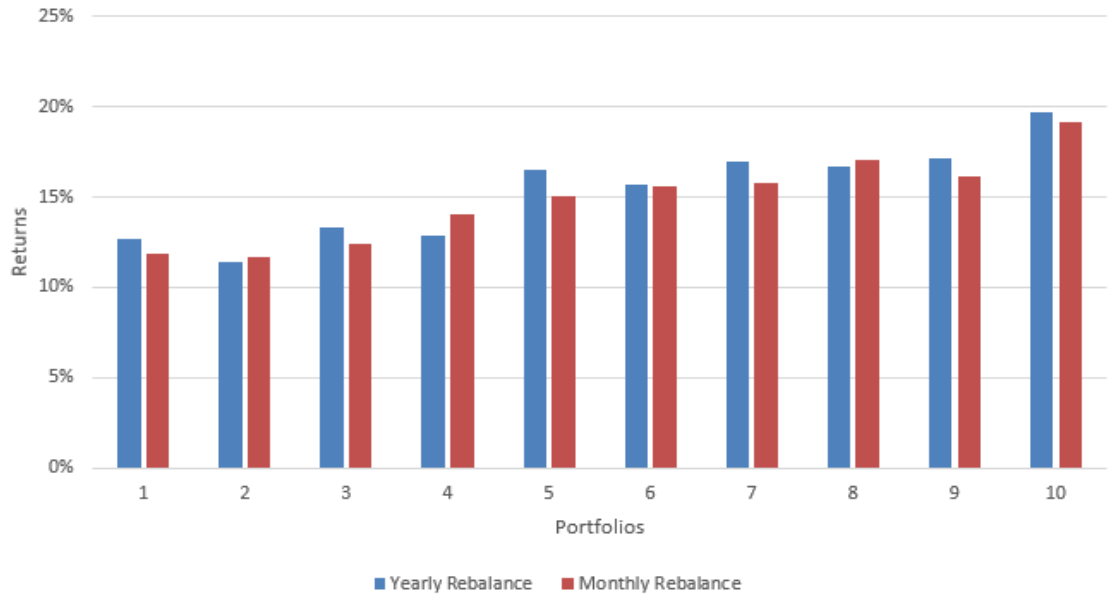
We used two variables: "Jobs Created" (number of job posting created by companies) and "Jobs Active" (number of job postings that are currently active and open). We also normalized these variables and transformed them using different change specifications. Then we deflated all variables by the firm market value of equity at the end of the year. In total, we had 10 variables.

We formed monthly and annual portfolios by dividing the sample of firms based on the 10 variables into both deciles and quintiles. The top portfolio (for deciles - portfolio 10) is the portfolio of 10% of the firms where the variable examined was the highest for a given month/year, and the bottom portfolio (portfolio 1) is the 10% of the firms where the variable is the lowest.

We then calculated each portfolio future returns after the formation period using both monthly and yearly return horizons. Finally, we calculated the hedge returns, i.e. the top portfolio average return minus the bottom portfolio average return (portfolio 10 minus portfolio 1 in the case of deciles).

The results suggest that there is alpha in the variables, with "Jobs Active" producing the highest and most consistent returns (Figure 15). Yearly hedge returns were between 6-8%. In the case of "Jobs Created", yearly hedge returns were between 2-4%.

Figure 15: Jobs Active Portfolio Returns



Source: Employment Data, Eagle Alpha

2. Equity > Quantitative > Consumer Transaction Data

Key Takeaway

Cross-sectional analysis of our partner dataset, performed by J.P. Morgan, generated annualised returns of 16.2% and a Sharpe ratio of 1.13.

Dataset

The dataset is delivered through a partnership with a provider that collects anonymized purchase data from around 2 million active shoppers, scanned from email purchase receipts. It covers over 600 merchants from more than 25 industries.

The provider transforms multiform unstructured email receipt data into a normalized digestible consumer transaction dataset. The dataset is also granular as it includes item and SKU-level transaction data, which is filtered into 53 product categories.

Geography	Coverage	Mapped to Tickers	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
USA	600 merchants	No	Since 2013	Real-Time	Weekly	7 days	Excel, TSV

Below we include an extract from J.P. Morgan's report "Big Data and AI Strategies. Machine Learning and Alternative Data Approach to Investing." J.P. Morgan analyzed a dataset of email receipts for 97 companies. 36 of these were private companies, and 61 public, 31 of which were S&P 500 constituents.

Taking liquidity into consideration, J.P. Morgan decided to test trading signals for the S&P 500 companies only.

Backtesting/Significance

We analyzed three time series: the dollar spend, number of orders and number of buyers. While number of orders and number of buyers are highly correlated (~99%), dollar spend is not highly correlated with number of buyers/orders (~25%).

We aggregated the daily spend/order/buyer data into a weekly score and calculated week-over-week percentage change for each. After winsorizing to 5th-95th percentile, we tested both the level and z-score as signals. Based on a cross-sectional comparison, we went long the top 5 stocks and short the bottom 5 stocks. The portfolio was rebalanced weekly.

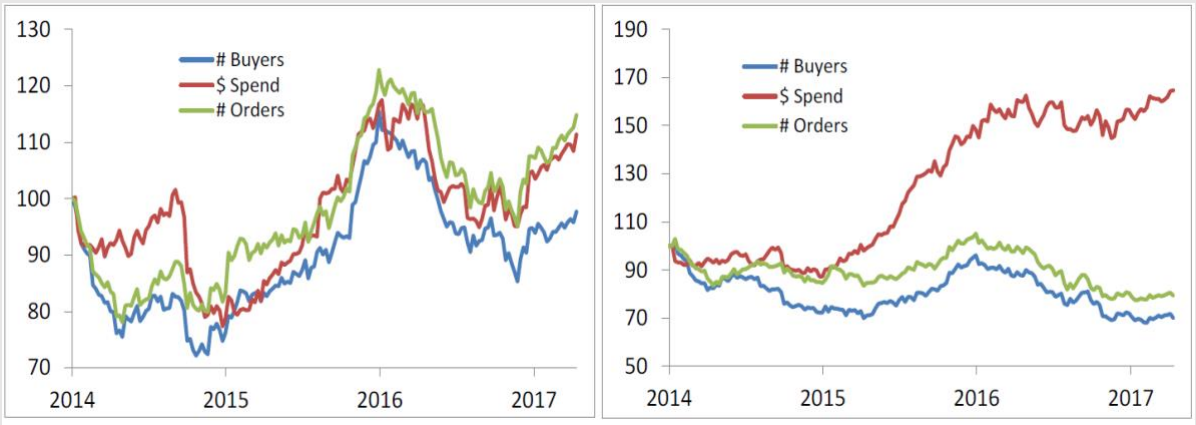
Figure 16: Sharpe ratios of various strategies using Dollar Spend, Buyer Count and Order Count

Dollar Spend Data	Top 6/ Bottom 6	Buyer Count Data	Top 6/ Bottom 6	Order Count Data	Top 6/ Bottom 6
Level	0.29	Level	0.02	Level	0.36
Z-score 4 week	1.13	Z-score 4 week	-0.71	Z-score 4 week	-0.49
Z-score 5 week	0.72	Z-score 5 week	-0.49	Z-score 5 week	-0.14
Z-score 6 week	0.67	Z-score 6 week	0.04	Z-score 6 week	0.11

Source: J.P. Morgan Macro QDS, Eagle Alpha

We also plot cumulative returns using the level (i.e. percentage of aggregated figure) and the 4-week z-score for all 3 datasets.

Figure 17: Performance of level (left) and time-series z-score of changes (right) as trading signals



Source: J.P. Morgan Macro QDS, Eagle Alpha

The 4-week z-score on the Dollar Spend metric displayed an impressive sharpe ratio of 1.13 (Figure 16). Annualised returns for this same portfolio also impressed at 16.2%. This can be seen in the right-hand chart in Figure 17 above.

3. Equity > Quantitative > Sentiment Data

Key Takeaway

Both Long and Short legs of Long/Short strategy created using sentiment data from this provider contributed to the global outperformance of the portfolio. The latter also exhibits a much lower maximum drawdown than the market due to the capacity of our explaining variable (Emotional Agitation) to anticipate market correction.

Dataset

This provider delivers investment signals based on sentiment and emotions analysis of over 120,000 data sources, focusing on social trading websites and specialized press. Signals are available for more than 2,000 tickers as well as bonds, currencies and commodities. History: since 2014.

Backtesting/Significance

For a weekly rebalancing frequency (λ=5), the provider simulated for each week an investment in a basket of n stocks (e.g., 30 for the DAX) where the position (Long or Short). A summary of the results for the DAX stocks is presented below. The grey line represents the DAX performance, the blue line shows net returns while the orange line shows gross returns.

Figure 18: Backtesting Results for DAX Stocks

Strategies (Daily Net Return)	Return Ann.	Volatility Ann.	Sharpe	Max DD
Long/Short	12.6%	18.7%	0.67	-18.2%
Long signal only	10.0%	10.2%	0.98	-13.8%
Short signal only	1.5%	15.1%	0.10	-18.2%
DAX	10.1%	19.6%	0.51	-28.3%

Source: Sentiment Data Provider

Disclaimer

This case study is directly from the vendor and Eagle Alpha has not backtested the data.

4. Equity > Quantitative > ESG Data

Key Takeaway

Backtesting shows that a portfolio optimized using ESG data from this data vendor outperforms both the benchmark index and a portfolio optimized using traditional factors.

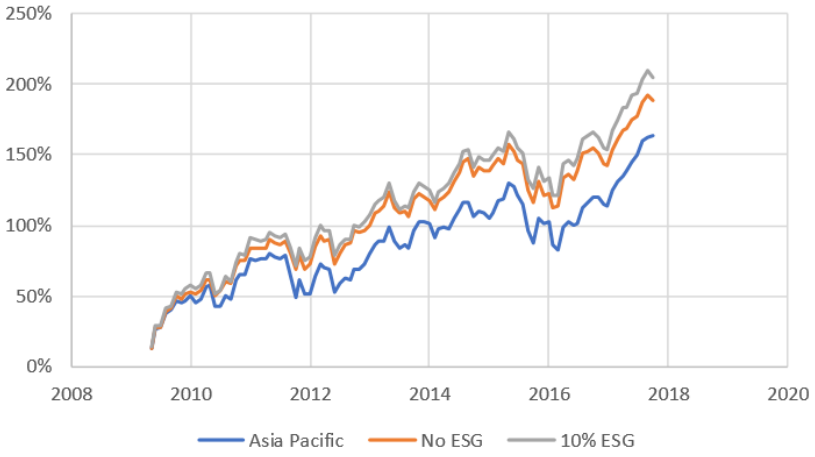
Dataset

This vendor provides a robust ESG (environmental, social, governance) dataset to institutional investors. They rate over 20,000 companies across a spectrum of ESG key performance indicators. Each KPI is represented by a score of 0-100 and can be used in various ways across investment strategies. Ratings begin as of 2009.

Case Study

Backtesting shows that a portfolio optimized using ESG data from this data vendor outperforms both the benchmark index and a portfolio optimized using traditional factors. This is true for both realized returns and shape ratios. The max drawdown for the ESG optimized portfolio is also less than the benchmark portfolios.

Figure 19: Asia Pacific Cumulative Returns



	Asia Pacific	No ESG	10% ESG
Realized Return (% Ann)	12.09%	13.30%	14.02%
Realized Risk (% Ann)	14.98%	13.38%	14.04%
Sharpe Ratio	0.81	0.99	1.00
Realized Active Return (% Ann)		1.21%	1.93%
Realized Active Risk (% Ann)		4.49%	3.97%
Information Ratio		0.27	0.49
Realized Beta	1.00	0.85	0.90
Max DrawDown	20.7%	17.6%	16.8%

Source: ESG Data Provider

5. Equity > Quantitative > ESG Data

Key Takeaway

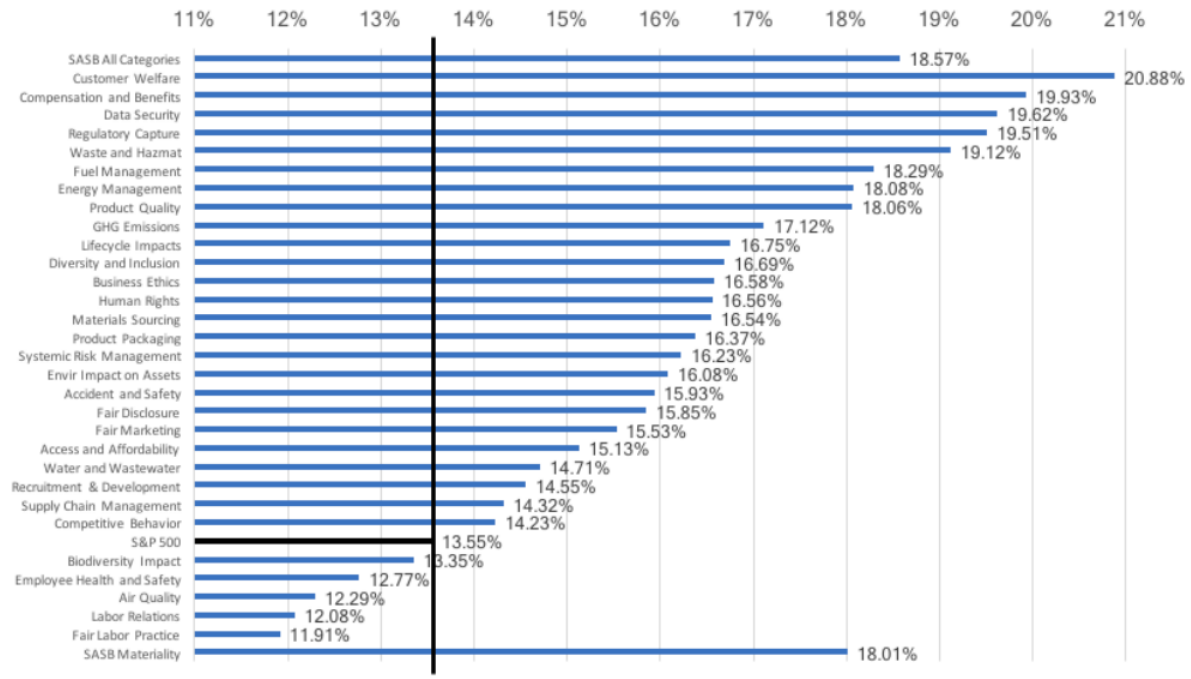
Backtesting shows that the provider's factors perform better than many traditional quant factors and are additive to a multi-factor investment approach.

Dataset

This vendor applies artificial intelligence (AI) to uncover timely ESG data on a variety of asset classes. The company delivers timely and investable insights by revealing intangible value and risk factors from unstructured text. The dataset history is September 2016.

The company adopted the framework designed by the Sustainability Accounting Standards Board (SASB). SASB has done extensive research on the sustainability categories that have material impact on companies in an industry. See Figure 20 below for more details.

Figure 20: Companies that Perform Well on the SASB Categories Outperform



Source: ESG Data Vendor

The SASB framework differentiates this vendor from other data providers. This vendor also performs daily company scores as events occur.

Case Study

The company scans tens of thousands of unstructured web sources and composes ESG scores for over 8,000 companies. Historical simulated backtesting results presented in Figure 21 show the total

return from a portfolio of stocks scored by the provider. These results do not include fees, transaction costs, commissions, taxes, or any other market frictions.

The provider's insight score is a moving average of the faster-moving event-based score while the momentum score is the relative ranking of the change in the insight score over the past 12 months.

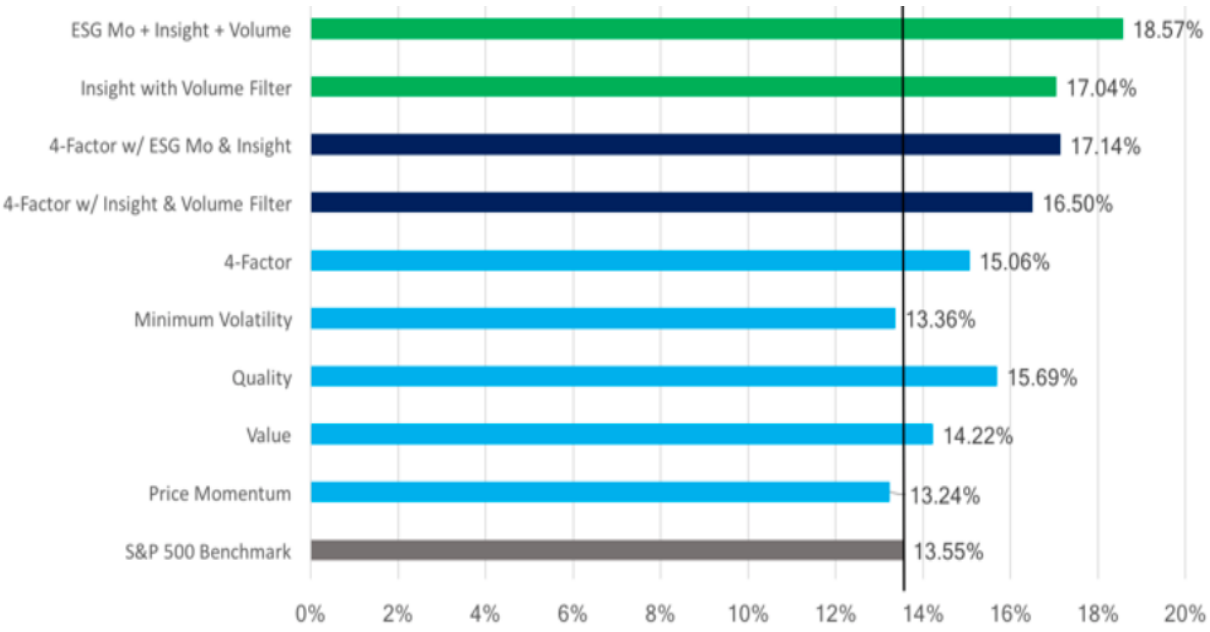
Figure 21: Cumulative Returns

		Cumulative Returns		Excess Returns	
Year	Benchmark	Insight	Momentum	Insight	Momentum
2013	27.2%	33.4%	32.6%	6.2%	5.4%
2014	14.0%	16.7%	20.1%	2.8%	6.1%
2015	-2.7%	3.2%	3.5%	5.9%	6.2%
2016	14.4%	12.3%	14.7%	-2.1%	0.3%
2017 YTD	9.4%	12.1%	13.6%	2.7%	4.2%
Since Inception	77.1%	103.0%	115.2%	25.9%	38.1%
Annualized	13.5%	17.0%	18.6%	3.5%	5.0%
Sharpe Ratio	1.00	1.29	1.36	0.29	0.36

Source: ESG Data Vendor

These scores provide a way to quantify many factors missed by traditional quantitative analyses. Backtesting shows that the provider's factors perform better than many traditional quant factors and are additive to a multi-factor investment approach.

Figure 22: Annualized Returns



Source: ESG Data Vendor

6. Equity > Quantitative > Employment Data

Key Takeaway

Testing showed that firms with lower employee turnover systematically outperformed those with higher turnover rates.

Dataset

This vendor offers a rich and unique view of the global labour force at a company, industry, and global level. The vendor maps half a billion individuals, more than 10,000 global public companies, and millions of private and non-corporate entities (government, education, military, healthcare, etc.) in order to capture workforce dynamics. History: since January 2007.

Case Study

Education and employment histories from an aggregator of public profile information for CRM system enrichment was used to gather information on more than 1.4 billion job change events (there are around 344 million profiles with basic details and 262 million profiles with employment records).

Testing showed that firms with lower employee turnover systematically outperformed those with higher turnover rates. In order to demonstrate this, the vendor computed monthly abnormal turnover rates (see the formula below), sorted firms by this characteristic and formed a long/short portfolio from the top and bottom quintiles.

$$AbnTurnover_{i,t} = Turnover_{i,t} - PredTurnover_{i,t},$$

Using a three-month lag, the long portfolio earned, on average, 1.23% monthly return, while the short portfolio earned a substantially lower return of 0.10% per month.

7. Equity > Discretionary (Short-Term) > Consumer Transaction Data

Key Takeaway

We have observed an accuracy of 80% across ten reports incorporating email receipt data published by Eagle Alpha's Data Analytics team. The signals also captured material stock moves on the day of earnings announcement.

Dataset

The dataset is delivered through a partnership with a provider that collects anonymized purchase data from around 2 million active shoppers, scanned from email purchase receipts. It covers over 600 merchants from more than 25 industries.

The provider transforms multi-form unstructured email receipt data into a normalized and digestible consumer transaction dataset. The dataset is also granular as it includes item and SKU-level transaction data, which is filtered into 53 product categories.

Geography	Coverage	Mapped to Tickers	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
USA	600 merchants	No	Since 2013	Real-Time	Weekly	7 days	Excel, TSV

Backtesting/Significance

Since the beginning of July 2017 Eagle Alpha's Data Analytics team published ten quarterly predictive reports utilizing email receipt data. In each instance Eagle Alpha built stock level predictive models using the email receipt data and then published reports highlighting whether our predictions suggested current consensus numbers appeared too high, too low or in-line. Below is an overview of the results.

Figure 23: Email Receipt Data Backtesting Results

Company	Report Date	Historic MAPE*	Eagle Alpha Prediction	Result	Stock Move on Earnings
Chipotle Mexican Grill	13-Jul-17	3.7%	Miss	Miss	-2.3%
PayPal	13-Jul-17	1.6%	Beat	Beat	2.3%
Square	13-Jul-17	1.4%	In-Line	Beat	-4.7%
Papa John's International	14-Jul-17	1.4%	Beat	Miss	9.6%
Starbucks	14-Jul-17	5.8%	Miss	Miss	-9.2%
GrubHub	16-Oct-17	2.6%	Beat	Beat	11.2%
Mindbody	17-Oct-17	1.1%	Beat	Beat	2.2%
Papa John's International	18-Oct-17	2.0%	In-Line	In-Line	-8.5%
PayPal	19-Oct-17	1.4%	Beat	Beat	5.5%
Square	19-Oct-17	2.4%	Beat	Beat	3.4%

* MAPE = Mean Absolute Percentage Error. It is the average of quarterly prediction errors for each predictive model.

Source: Eagle Alpha Analysis, Email Receipt Data

Of the ten reports, eight proved accurate, indicating a hit rate of 80%. The table above also shows there was significant alpha in the signals, as indicated by the stock move on the day of the earnings.

8. Equity (Just Eat JE) > Discretionary (Short-Term) > Consumer Transaction Data

Key Takeaway

Using consumer transaction data, Eagle Alpha's predictive model for Just Eat correctly pointed to stronger than expected revenues in H2 2015 and H1 2016.

Dataset

Consumer transaction data is provided by Eagle Alpha's partner and consists of anonymized panel data drawn from users of the UK's largest personal financial assistant, an app/web platform. Data is collected in near real-time, directly from bank and credit card records.

The panel consists of 230 million transactions across over 500,000 accounts and grows by an average of 5 million transactions per month. With each panel member holding an average of 4 financial accounts, across multiple providers, this consumer transaction dataset represents the UK's largest complete collection of consumer financial activity.

Geography	Coverage	Mapped to Tickers	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
UK	U.K. National, with regional breakdowns	Yes	Since 2012	Daily	Daily, Weekly, Fortnightly & Monthly	3 days	AWS S3, FTP

Backtesting/Significance

Eagle Alpha's Just Eat model based on U.K. consumer transaction data has shown a MAPE of 1.8% in backtesting, a large improvement on consensus which has an MAPE of 3.7%.

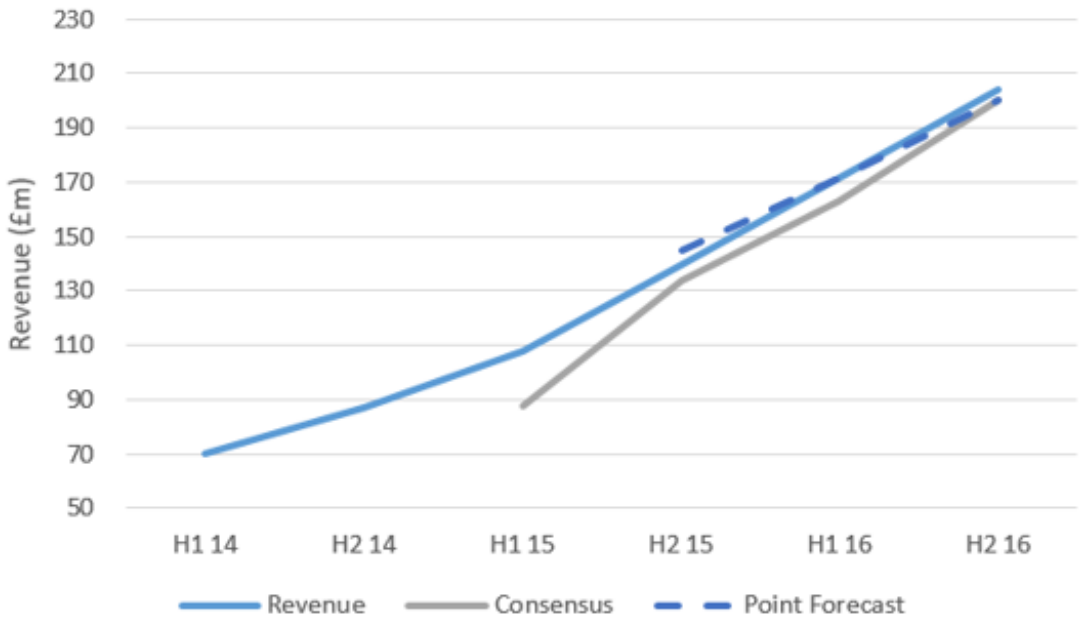
Note: testing of this dataset across 41 UK retailers found that it improved the predictive power of a baseline ARIMA model in 88% of cases.

Case Study

Using consumer transaction data, Eagle Alpha's predictive model for Just Eat correctly pointed to stronger than expected revenues in H2 2015 and H1 2016.

Our ARIMAX model forecasted H1 2016 revenues at 171.6m compared to a consensus estimate of 163m. In H2 2015 our model predicted revenues of 144.8m compared to a consensus estimate of 134m.

Figure 24: Eagle Alpha's Model Predicted Stronger Than Expected Results



Source: Eagle Alpha Analysis, Consumer Transaction Data, Bloomberg

Actual revenues for the second half of 2015 came in at 140m beating consensus, in line with our prediction. Following the H1 2016 beat the Just Eat share price rose by 4.3% on the day and 17.6% on the month. Similarly, the stock rose by 3% after the H2 2015 beat.

9. Equity (Starbucks SBUX) > Discretionary (Short-Term) > Consumer Transaction

Key Takeaway

Eagle Alpha's predictive model for Starbucks using email receipt data accurately predicted a revenue beat in Q2 2017.

Datasets

The dataset is delivered through a partnership with a provider that collects anonymized purchase data from around 2 million active shoppers, scanned from email purchase receipts. It covers over 600 merchants from more than 25 industries.

The provider transforms multi-form unstructured email receipt data into a normalized and digestible consumer transaction dataset. The dataset is also granular as it includes item and SKU-level transaction data, which is filtered into 53 product categories.

Geography	Coverage	Mapped to Tickers	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
USA	600 merchants	No	Since 2013	Real-Time	Weekly	7 days	Excel, TSV

Google Trends is a public web facility based on Google Search that shows how often a particular search term is entered relative to the total search-volume over time across various regions of the world. Using Google Trends, Eagle Alpha has built company specific indices based on search terms that are related to a given retailer's product offering. This involves an exhaustive process for identifying search terms related to a company's revenues using both internal and third party tools.

Backtesting/Significance

Email Receipt Data

Our Starbucks model, based on email receipt data, has a MAPE of 5.8% and the standard deviation of our error is 6.9%. Note: broader testing of the email receipt dataset across 66 US retailers found that it improved the predictive power of a baseline ARIMA model in 60% of cases.

Over the last 19 quarters our Starbucks search index based on Google search volumes has demonstrated a two-quarter hit rate of 68% in in-sample testing, and a three quarter hit rate of 53%. This means that 68% of the time the three-month moving average has moved in the same direction as same store sales over a two-quarter period.

Online Search Data

Eagle Alpha's Data Analytics team has published 36 quarterly reports for consumer companies incorporating Google search data. 25 of these indicators proved accurate equating to a hit rate of 69%.

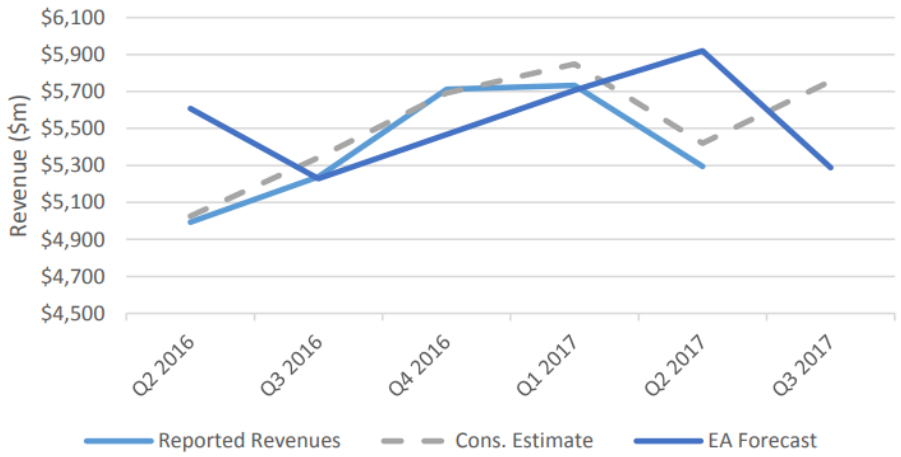
Case Study

Using email receipt and Google search data, Eagle Alpha published a research report on July 14th 2017 predicting a weaker June quarter for Starbucks versus consensus.

Figure 25 below shows that according to Eagle Alpha's predictive model Bloomberg estimates for Starbucks FYQ3 were too high. Consensus revenue estimates were at $5.75bn as analysts were predicting a YoY revenue growth of 9.9%.

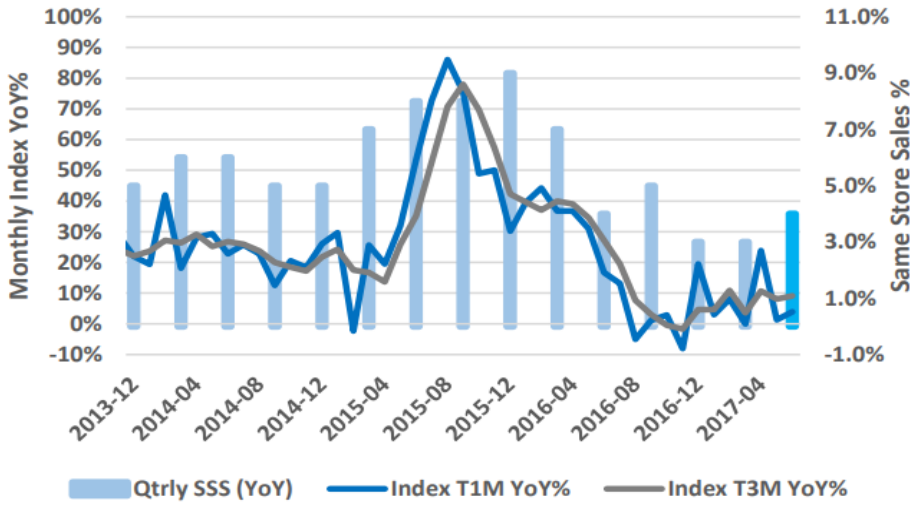
Our Starbucks search index (Figure 26) was in line with the previous two quarters, calling into question consensus expectations for SSS to accelerate later in 2017.

Figure 25: Eagle Alpha Model Predicted June Quarter Downside for Starbucks



Source: Eagle Alpha Analysis, Email Receipt Data, Bloomberg

Figure 26: Search Data Suggested Stagnant Quarterly Growth



Source: Eagle Alpha Analysis, Email Receipt Data, Bloomberg

On July 28th 2017, Starbucks reported weaker-than-expected results with revenues of $5.66bn versus consensus estimates of $5.75bn. The company lowered its full year 2017 forecast giving weak same store sales guidance. Starbucks also decided to close all 379 of its Teavana stores.

10. Equity (Great Wall Motors 601633 CH) > Discretionary (Short-Term) > Pricing Data

Key Takeaway

In Q4 2014, The CAI (China Auto Insight) data was more accurate than consensus estimates at predicting an important inflection in revenue growth.

Dataset

Eagle Alpha's CAI dataset is provided through a partnership agreement with a leading Chinese financial automotive consultant. The dataset is collected using a large panel of dealerships throughout China, combining other data sources such as web data and more traditional data sets to create a large and well-structured database.

Geography	Coverage	Mapped to Tickers	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
China	National, with regional breakdowns	No	Since 2012	Mixed – Month/ Bi-Monthly	Mixed – Month/ Bi-Monthly	Between 5 & 20 days	API, CSV

Backtesting/Significance

The CAI data shows a 99% correlation with revenues for domestic Chinese auto manufacturer Great Wall Motors (601633 CH) over a 5-year period, and a 95% correlation with YoY revenue growth. The calculated dealership revenue also correctly projected the directional movement of reported revenues for Great Wall in fourteen of the fifteen quarters between Q1 2013 and Q3 2016.

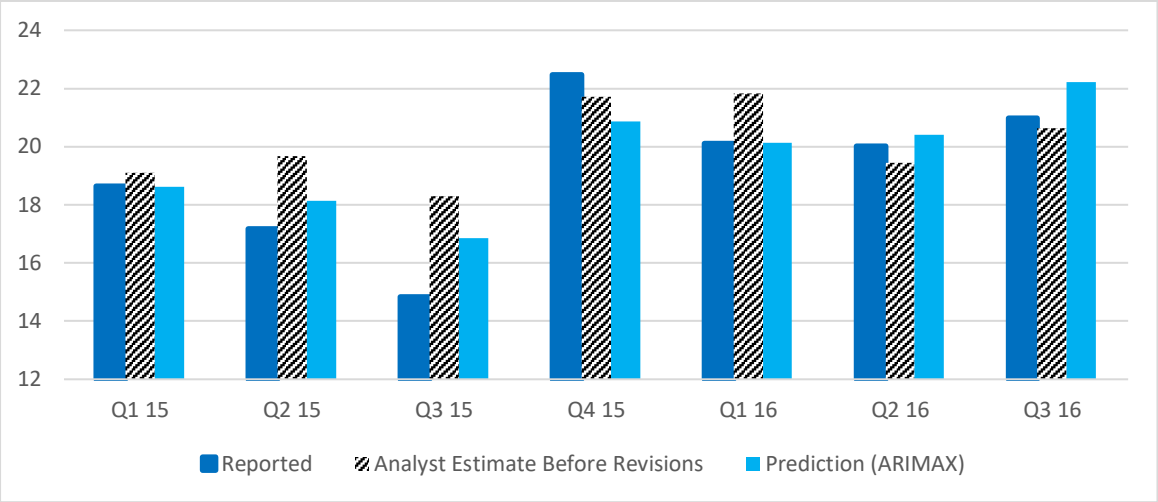
An ARIMAX[24] model for predicting Great Wall Motors (601633 CH) revenue demonstrated an out-of-sample MAPE of 4.9%. This compares to a MAPE of 10.3% on a baseline ARIMA model using just historic revenues. The error rate for market consensus estimates was 8.1% over the same period. Directional accuracy is also markedly improved over the baseline model, increasing from 57% to 86%.

Case Study

The 4.9% error rate discussed above is better than the error rate for market consensus estimates of 8.1% over the same period. This is shown in Figure 27 below. The analyst estimates in this calculation were taken one month after the end of the quarter, as the CAI data is published between 5 and 20 days after month-end. It's clear that the predictive model using CAI data provides investors with a more accurate revenue estimate for Great Wall Motors.

[24] Autoregressive integrated moving average with exogenous variables

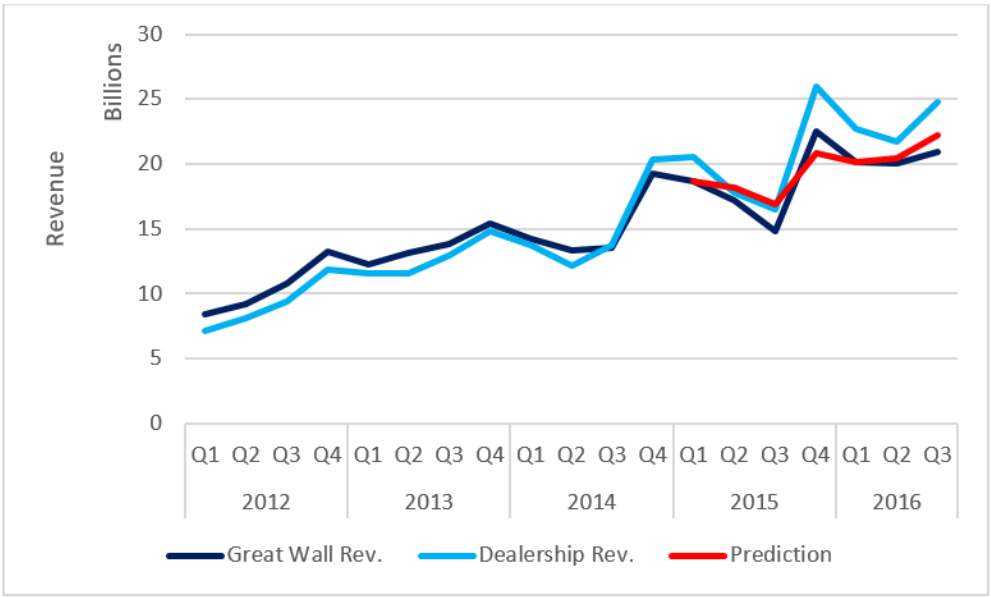
Figure 27: Reported Numbers vs Estimates vs Eagle Alpha Prediction



Source: Eagle Alpha Analysis, Bloomberg, CAI Data

In Q4 2014, the CAI data was more accurate than consensus estimates at predicting an important inflection in revenue growth. Analyst estimates were for QoQ revenue growth of 17% for Great Wall Motors, compared to reported QoQ growth of 42%. Estimated QoQ growth based on the CAI raw data was much closer at 48%.

Figure 28: Great Wall Revenue Prediction



Source: Eagle Alpha Analysis, Bloomberg, CAI Data

11. Equity (AMZN) > Discretionary (Short-Term) > Location Data

Key Takeaway

Foot traffic data to Whole Foods was used to track price reductions.

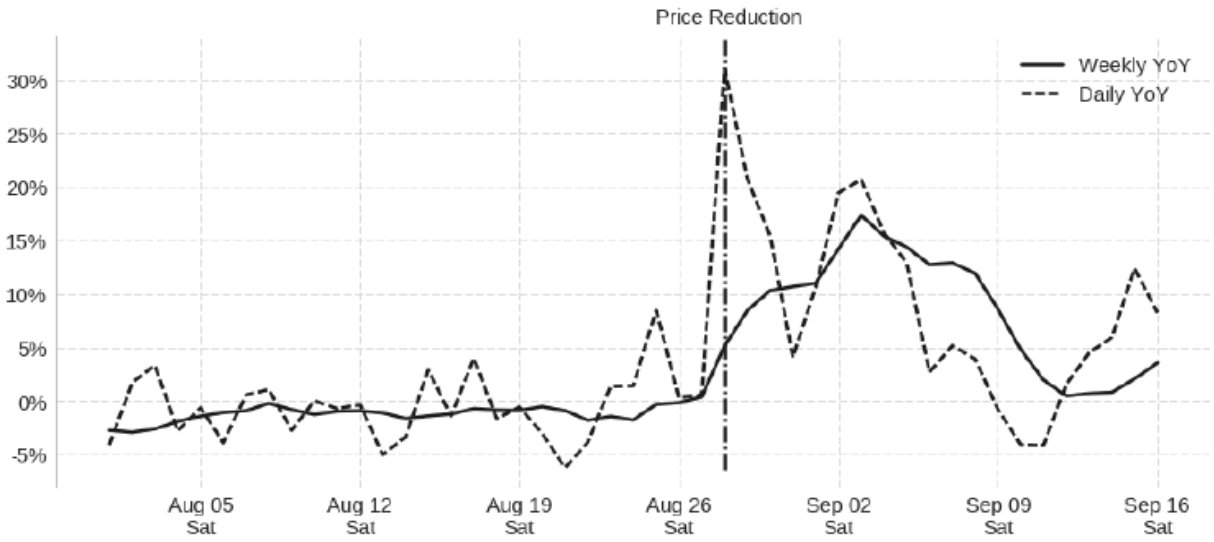
Dataset

The provider transforms real-time locations from mobile phones across the globe into objective and actionable insights on the performance of businesses, markets, and economies.

Case Study

In order to measure the change in foot traffic to Whole Foods in response to the price reduction on August 28, while also controlling for the impact of seasonality and Labor Day weekend, The figure below shows the year-over-year ("YoY") change for daily and weekly rolling windows of foot traffic. For the week ending Sunday, September 3, the weekly YoY change in foot traffic to Whole Foods peaked at 17%. On August 28, the daily YoY change in foot traffic peaked at 31%.

Figure 29: YoY Change in Foot Traffic for Whole Foods' Customers



Note: The weekly rolling window is defined as a seven-day period of time that increments or rolls forward by adding one new day and simultaneously dropping the oldest day. The rolling window is plotted in terms of the YoY change relative to the comparable seven-day period in 2016.

Source: Location Data Provider

12. Equity (Dick's Sporting Goods DKS) > Discretionary (Short-Term) > Satellite Imagery

Key Takeaway

Satellite data analysis showed falling traffic at Dick's Sporting Goods since the start of 2017. On August 15th 2017, Dick's Sporting Goods reported worse-than-expected results.

Dataset

The case study presented below was prepared by a satellite imagery data analytics company that catalogued and backtested more than one million parking lot images that accounted for 1.5 billion cars over seven years. The traffic data obtained from satellite images can then be used to analyze cumulative YoY car count growth rates and compare it to revenue growth and share price dynamics.

Case Study

Car counts at Dick's Sporting Goods have been falling since the start of 2017. There was a 7.2% YoY drop in Q1 2017 and then a further 8.4% YoY drop in Q2 2017 (Figure 30). The downward trend continued in July and August 2017 proving to be an early indicator of Q2 2017 results.

On August 15th 2017, Dick's Sporting Goods reported worse-than-expected results with same store sales rising by 0.1% which was much lower than the company's forecast of 2-3% and the consensus estimate of 1.7%.

Figure 30: DKS Cumulative YoY% Change in Car Counts

Source: Satellite Data Analytics Provider

Disclaimer

This case study is directly from the vendor and Eagle Alpha has not backtested the data.

13. Equity (Chipotle CMG) > Discretionary (Short-Term) > Satellite Imagery

Key Takeaway

Satellite data analysis showed falling traffic at Chipotle Q4 2014. This analysis was ahead of the street as sell side analysts only began revising Chipotle down 3 quarters later.

Dataset

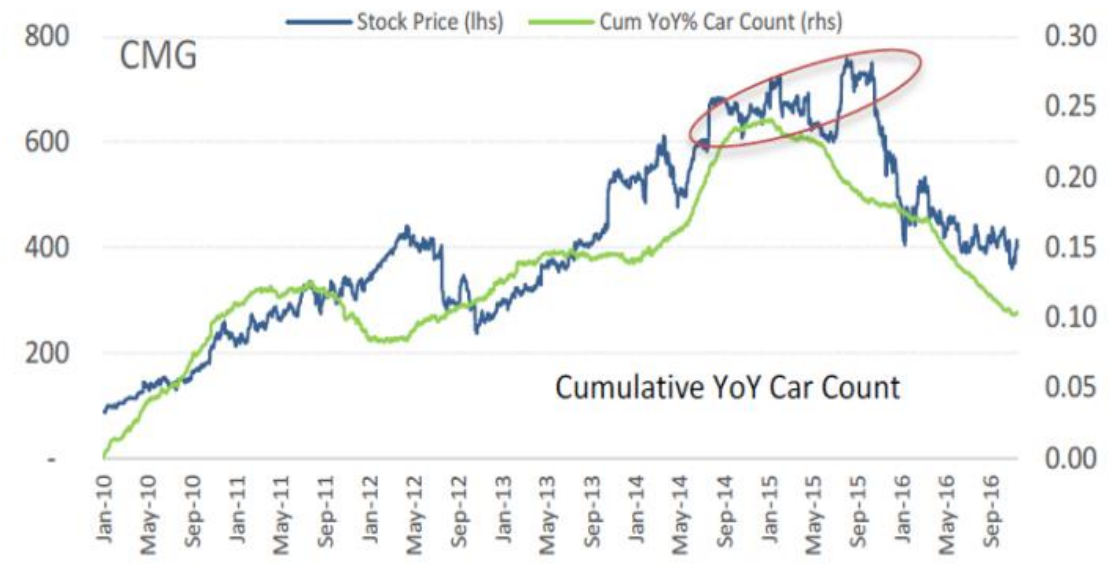
The case study presented below was prepared by a satellite imagery data analytics company that catalogued and backtested more than one million parking lot images that accounted for 1.5 billion cars over seven years. The traffic data obtained from satellite images can then be used to analyze cumulative YoY car count growth rates and compare it to revenue growth and share price dynamics.

Case Study

Figure 31 below shows how satellite imagery data was used to study Chipotle's structural inflection points over the 2014-2016 period. The car count traffic peaked in late 2014 before recording a quick fall from then onwards. This analysis was ahead of the street as sell side analysts only began revising Chipotle down 3 quarters later.

The E. coli outbreak in October 2015 had a big impact on the company's share price and the following outbreak of norovirus in March 2016 caught investors' attention as well. Most recently, Chipotle reported lower than expected Q4 2016 results on 2 February 2017.

Figure 31: CMG Cumulative Car Count vs Stock Price



Source: Satellite Data Analytics Provider

Disclaimer

This case study is directly from the vendor and Eagle Alpha has not backtested the data.

14. Equity (CMG) > Discretionary (Short-Term) > Employment, Online Search

Key Takeaway

Negative growth in active job listings and Google search data were used to correctly call worsening momentum for Chipotle.

Datasets

Job Listings Data

Completely unique in the industry, the job listing dataset only indexes jobs directly from employer websites. Updated daily with over 4 million jobs from more than 30,000 employers, the platform eliminates duplicate and expired job listings, as well as job pollution.

Google Trends Data

Google Trends is a public web facility based on Google Search that shows how often a particular search-term is entered relative to the total search-volume over time across various regions of the world.

Backtesting/Significance

Job Listings Data

Backtesting of the dataset demonstrated alpha in the dataset, with “Jobs Active” producing the highest and most consistent returns. Yearly hedge returns were between 6-8%.

Google Trends Data

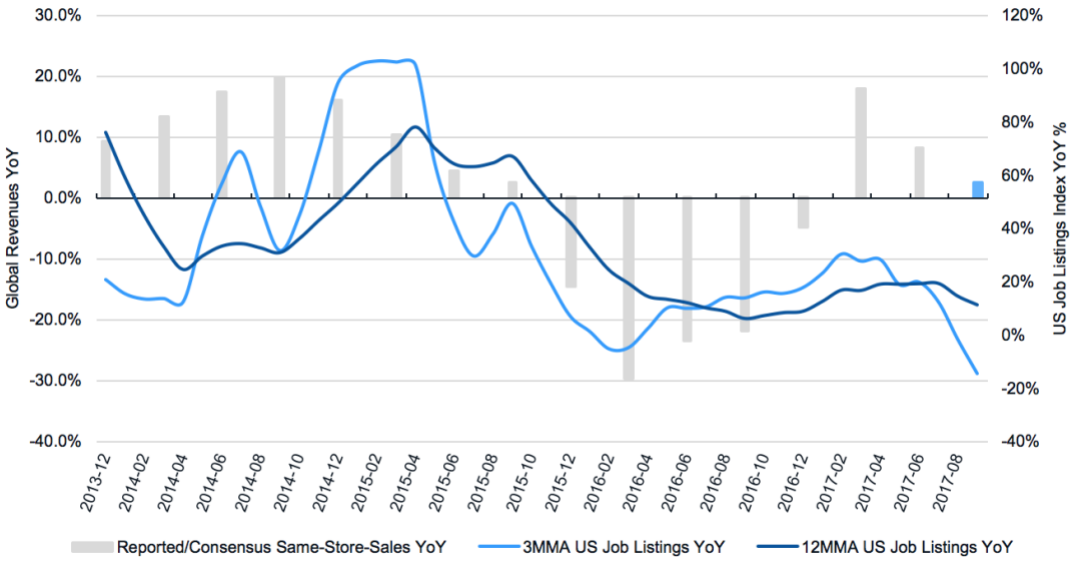
Over the last 20 quarters the Chipotle indicator has demonstrated a three-quarter hit rate of 90% in in-sample testing. This means that 90% of the time the three-month moving average has moved in the same direction as same store sales over a three-quarter period. One-quarter hit rate is 82%.

Case Study

On 16th October 2017, Eagle Alpha published a Data Analytics report titled “Chipotle: Job Listings and Online Search Point to Further Slowdown”. Figure 32 below shows 3 and 12 month moving averages of company job listings data against quarterly same store sales (SSS).

Growth in active job listings has been a strong indicator of inflections in revenue growth in the past (e.g. SSS slowdown in 2015 and bottom out in mid-2016). Job listings growth inflected down in Q2 2017 and the data was showing us that the longer-term momentum has stalled.

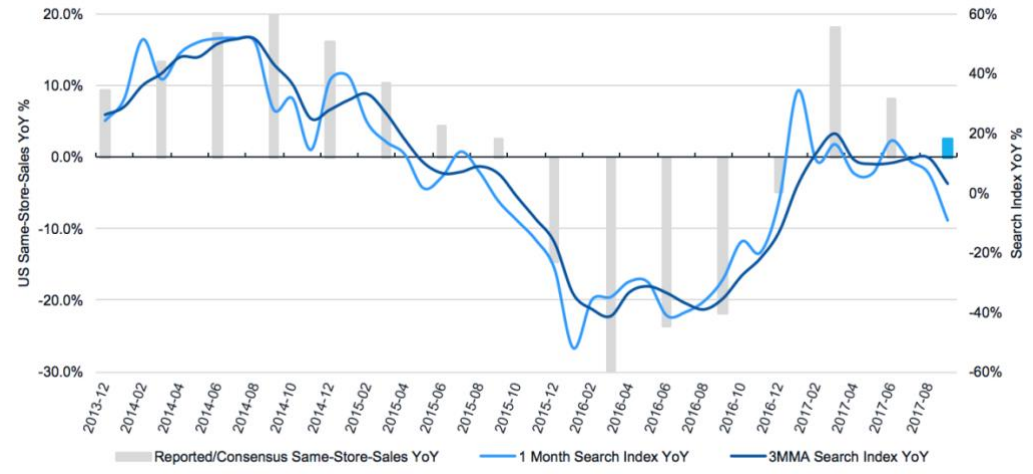
Figure 32: Chipotle Job Listings Index



Source: Job Listings Data

Figure 33 below shows that our Search Signal indicator, built using Chipotle specific terms from Google Trends, flattened in the June quarter and then turned negative as the 1 month index crossed below the 3 MMA. In our report, we noted that "historically there has been about a one quarter lead time before search terms get reflected in reported numbers".

Figure 33: Chipotle Search Signal



Source: Job Listings Data

On October 25th 2017, Chipotle reported weak Q3 numbers and lowered guidance for the rest of the year.

15. Equity (AR) > Discretionary (Short-Term) > Sensor Data

Key Takeaway

The sensor data provider observed production increases at Antero Resources and correctly anticipated management would raise company guidance.

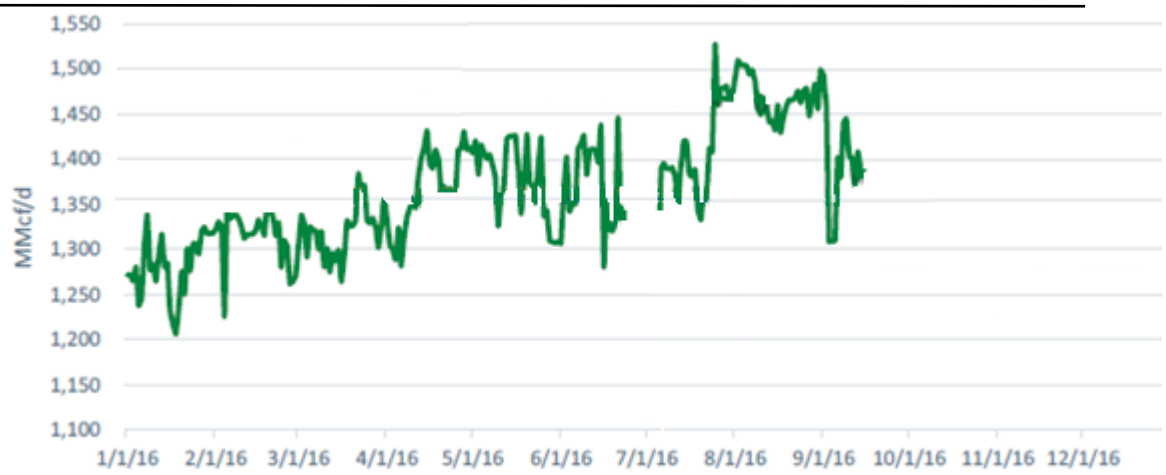
Dataset

The vendor provides energy market data and intelligence across the oil, power, natural gas and LNG, agriculture, petrochemical and NGLs. The company measures market fundamentals using thousands of patented and proprietary land, sea, and satellite monitors strategically deployed worldwide, delivering exceptional insight and intelligence to clients. History: since January 1999.

Case Study

In 2016, the vendor monitored Antero Resources' seven rigs: six in Marcellus, NY and one in Utica, NY. The sensor data provider observed production increases and anticipated management to raise the company guidance.

Figure 34: Antero Resources Daily Productions in 2016



Source: Sensor Data Provider

In September 2016, Antero Resources announced: "The increase in production guidance… is primarily a function of the improved recoveries and drilling efficiencies Antero has achieved throughout the year."

16. Equity (RSH) > Discretionary (Short-Term) > Geo-Location

Key Takeaway

Geo-location data was used to anticipate disappointing sales at RadioShack stores.

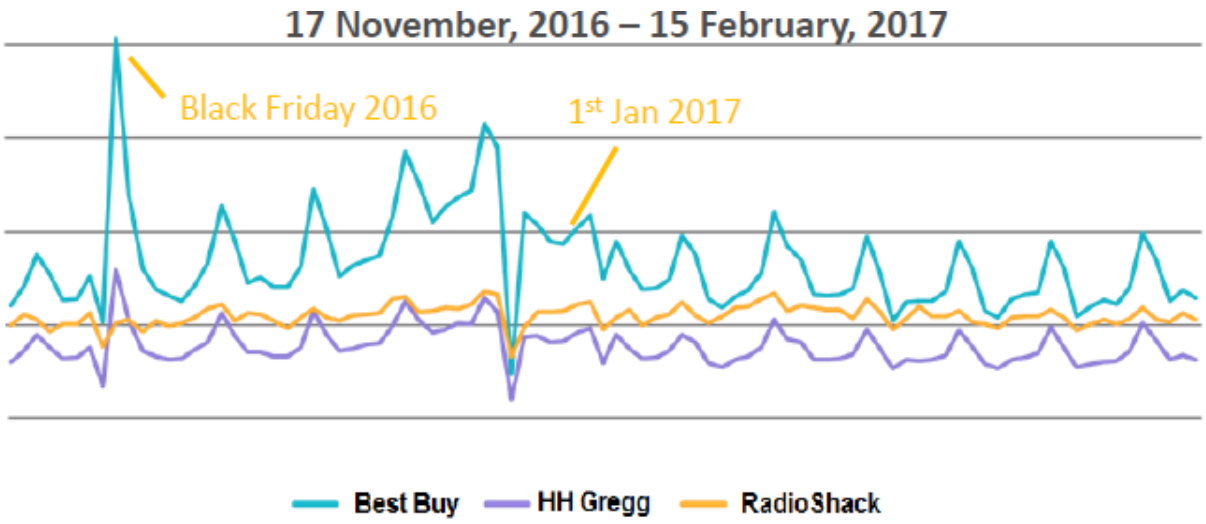
Dataset

The vendor receives raw location signals from 100s of app partners and also collects data from its own apps. In total it processes 1.7 billion visits per month from 130 million devices globally. History: since April 2016.

Case Study

The vendor analysed real-time tracking data in the 2016 holiday period. The figure below shows that RadioShack's sales remained more or less stable throughout the period, unlike major competitor Best Buy.

Figure 35: Visitors Per Store



Source: Geo-Location Data Provider

On March 9th 2017, RadioShack announced bankruptcy and the closure of more than 200 stores citing decreased holiday sales as a contributing factor.

17. Equity (GoPro GPRO) > Discretionary (Shorter-Term) > Pricing Data

Key Takeaway

Online pricing data pointed to negative GoPro (GPRO) fundamentals, which were reflected in GoPro's subsequent results and, ultimately, GoPro's stock price.

Dataset

Pricing data is crawled from the websites of large retailers. This can provide an insight into long term trends, as well as the most recent trading performance, and is based on metrics such as average selling price and share of bestselling products in a category. Eagle Alpha owns this dataset.

The dataset currently supports over 100 brands/companies from categories including Tech Hardware (e.g. Apple, Cisco, Canon), Consumer Electronics (e.g. Sony, LG Electronics, Harman), Household Appliances (e.g. Whirlpool, Electrolux) and Leisure Products (e.g. Mattel, Hasbro) and the list is growing all the time.

Geography	Coverage	Mapped to Tickers	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
U.S	U.S Consumer Products Market	Yes	Since 2013	Daily, Weekly and Monthly	Daily, Weekly and Monthly	1 day	Excel/CSV

Backtesting/Significance

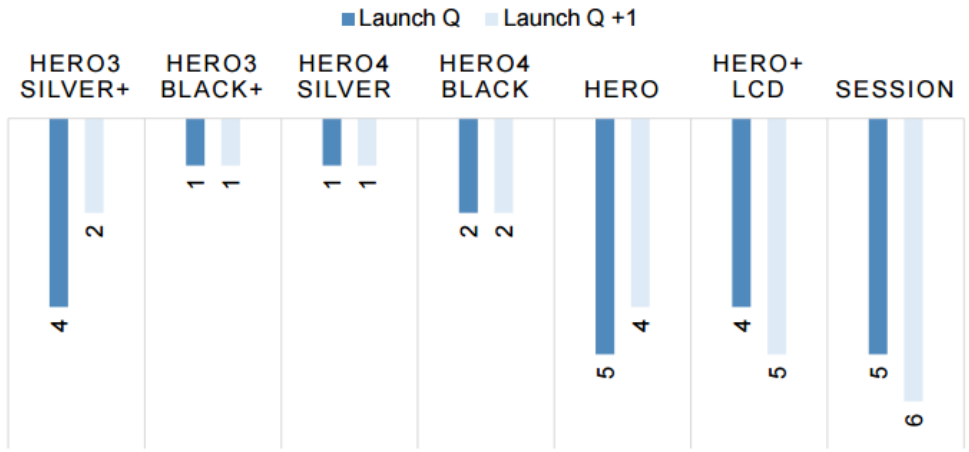
Pricing data has proved valuable at providing a directional indicator for the sales of consumer products, particularly around major product launches.

Note: Eagle Alpha has published 21 quarterly reports on consumer stocks incorporating crawled pricing data. 13 of these indicators proved accurate equating to a hit rate of 62%.

Case Study

Eagle Alpha first published a Data Analytics report on action camera manufacturer GoPro (GPRO US) on October 21st 2015. The data from US electronics websites pointed to potential weakness in GoPro revenue for the third quarter of that year. The crawled data was showing weak demand for GoPro's products, and a negative mix shift to lower end products that was likely to impact average selling prices (ASP). The report also highlighted weakness in the ranking of bestselling cameras, including the Session product which had recently been released (Figure 36).

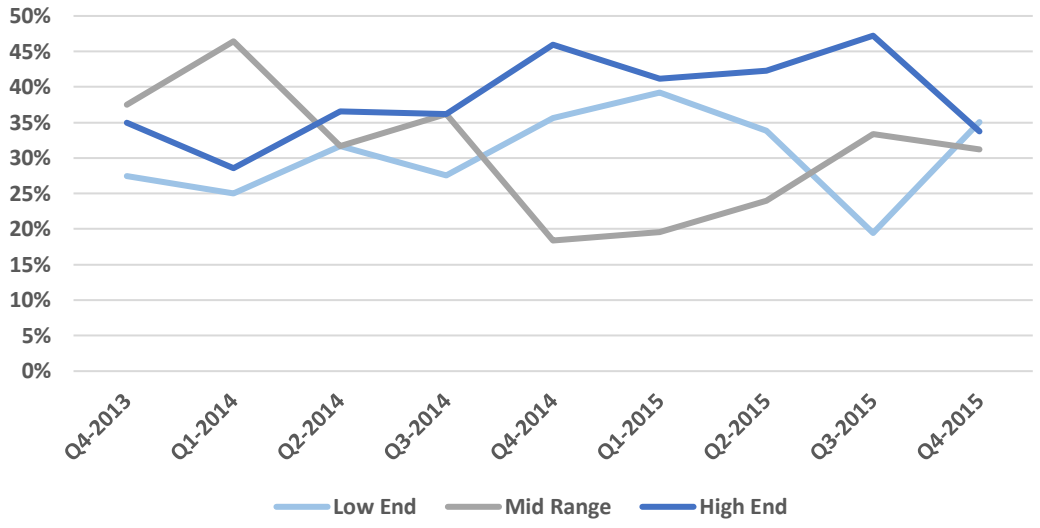
Figure 36: Weakness in the Ranking of Bestselling Cameras



Source: Eagle Alpha Analysis, E-Commerce Data

Subsequent Eagle Alpha reports in 2016 continued to point out that demand for GoPro was diminishing and that ASP trends continued to remain under pressure. A report on February 9th 2016 titled "Focus on Entry Level Products Ignores Issues Elsewhere" showed that GoPro was seeing price pressure across all price points and that growth in lower end cameras was unlikely to compensate for price and share compression of mid-range and higher end products (Figure 37).

Figure 37: Split of Bestsellers by Price Segment

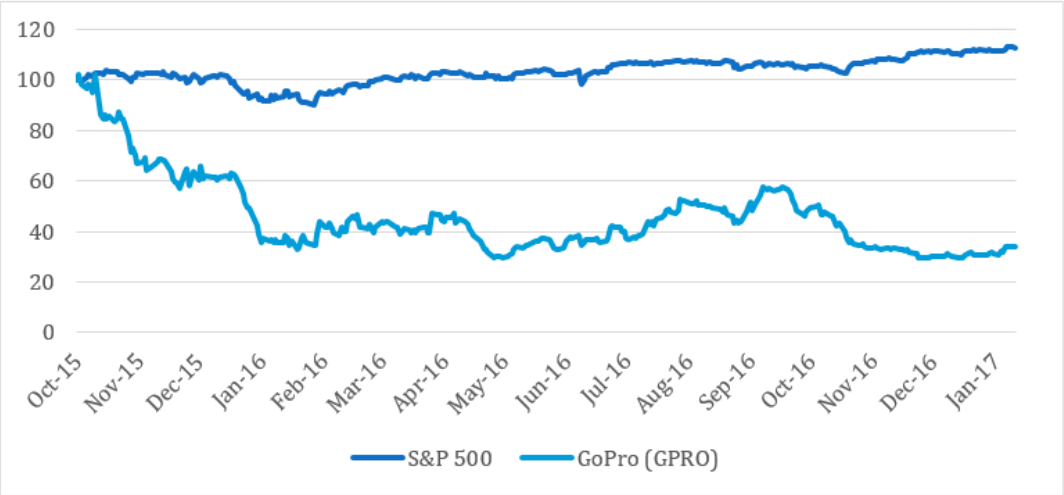


Source: Eagle Alpha Analysis, E-Commerce Data

By the September quarter of 2016 Eagle Alpha analysis was showing that "key elements are stabilizing for the action camera brand … but YoY pressure persists".

These negative fundamentals were reflected in GoPro's stock price over the period of Eagle Alpha's coverage. The relative performance of GoPro compared to the S&P 500 over the period can be seen in the following chart (Figure 38).

Figure 38: Indexed Share Price Performance



Source: Eagle Alpha Analysis, Bloomberg

18. Equity (Fitbit FIT) > Discretionary (Longer-Term) > Pricing Data

Key Takeaway

Online retail data showed improving sell-through trends for Fitbit in the first half of 2017. On August 2nd 2017, Fitbit reported better than expected results with adjusted revenue of $353.3m vs consensus estimate of $339.2m.

Dataset

Pricing data is crawled from the websites of large retailers. This can provide an insight into long term trends, as well as the most recent trading performance, and is based on metrics such as average selling price and share of bestselling products in a category. Eagle Alpha owns this dataset.

The dataset currently supports over 100 brands/companies from categories including Tech Hardware (e.g. Apple, Cisco, Canon), Consumer Electronics (e.g. Sony, LG Electronics, Harman), Household Appliances (e.g. Whirlpool, Electrolux) and Leisure Products (e.g. Mattel, Hasbro) and the list is growing all the time.

Geography	Coverage	Mapped to Tickers	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
U.S	U.S Consumer Products Market	Yes	Since 2013	Daily, Weekly and Monthly	Daily, Weekly and Monthly	1 day	Excel/CSV

Backtesting/Significance

Eagle Alpha has published 21 quarterly reports on consumer stocks incorporating crawled pricing data. 13 of these indicators proved accurate equating to a hit rate of 62%.

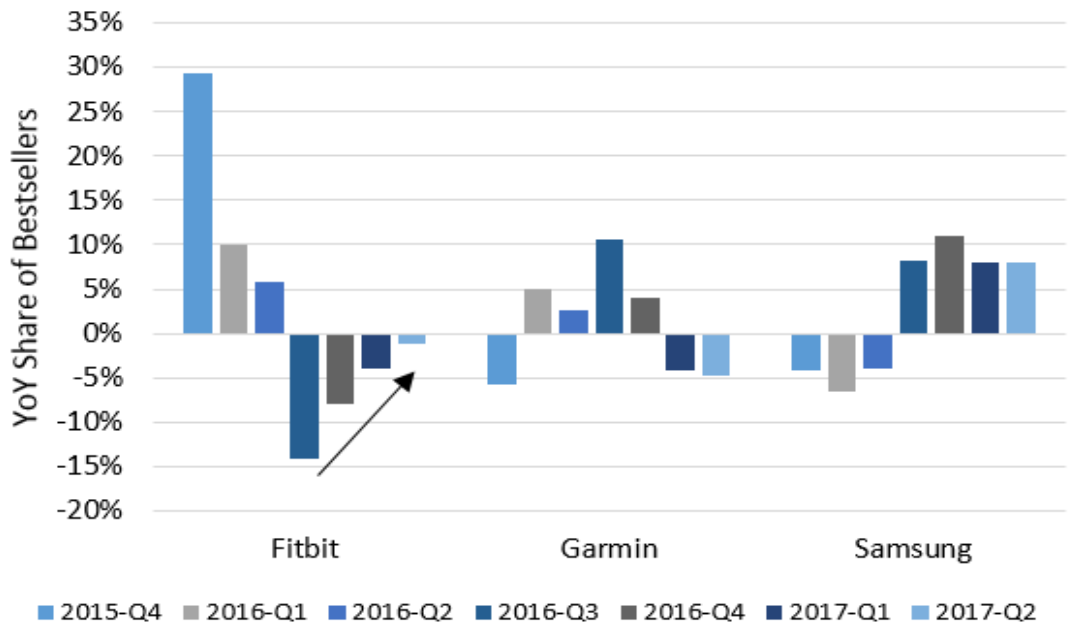
Case Study

On July 4th 2017, Eagle Alpha published a research note analyzing Fitbit's Q2 2017 with the use of online retail pricing data. Our analysis showed improving sell-through trends in the first half of the year. We noted the company's previous comments on improving inventory and concluded that Fitbit would report stronger than expected revenue.

Figure 39 below shows that Fitbit's share of bestsellers stabilized in Q2 as it further consolidated its position as the number one ranked fitness watch (Figure 40).

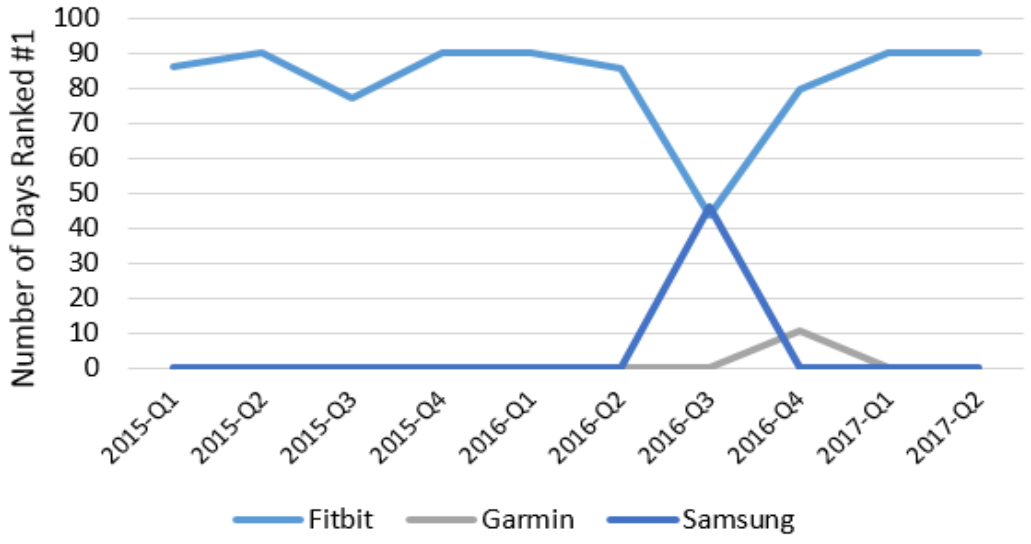
On August 2nd 2017, Fitbit reported better than expected results with adjusted revenue of $353.3m vs consensus estimate of $339.2m. Inventory was again mentioned by CEO James Park following these results. James stated "consumer demand in the second quarter was better than anticipated, enabling Fitbit to reduce channel inventory and generate better sales," which was in line with our analysis.

Figure 39: Online Data Showed Fitbit's Improving Share Trends



Source: Eagle Alpha Analysis, Online Retail Data

Figure 40: Fitbit Consolidated Top Ranking in Fitness Watches in Q2



Source: Eagle Alpha Analysis, Online Retail Data

19. Equity (HubSpot HUBS) > Discretionary (Long-Term) > Pricing Data

Key Takeaway

Accelerating growth in active job listings and Google search data were used to correctly call improving momentum for HubSpot.

Datasets

Job Listings Data

Completely unique in the industry, the job listing dataset only indexes jobs directly from employer websites. Updated daily with over 4 million jobs from more than 30,000 employers, the platform eliminates duplicate and expired job listings, as well as job pollution. From the core platform, the company has developed an array of products and services and achieved significant market traction in two primary business units: Candidate Sourcing and Job Market Data and Analytics.

Google Trends Data

Google Trends is a public web facility based on Google Search that shows how often a particular search-term is entered relative to the total search-volume over time across various regions of the world.

Backtesting/Significance

Job Listings Data

Backtesting of the dataset demonstrated alpha in the dataset, with "Jobs Active" producing the highest and most consistent returns. Yearly hedge returns were between 6-8%.

Google Trends Data

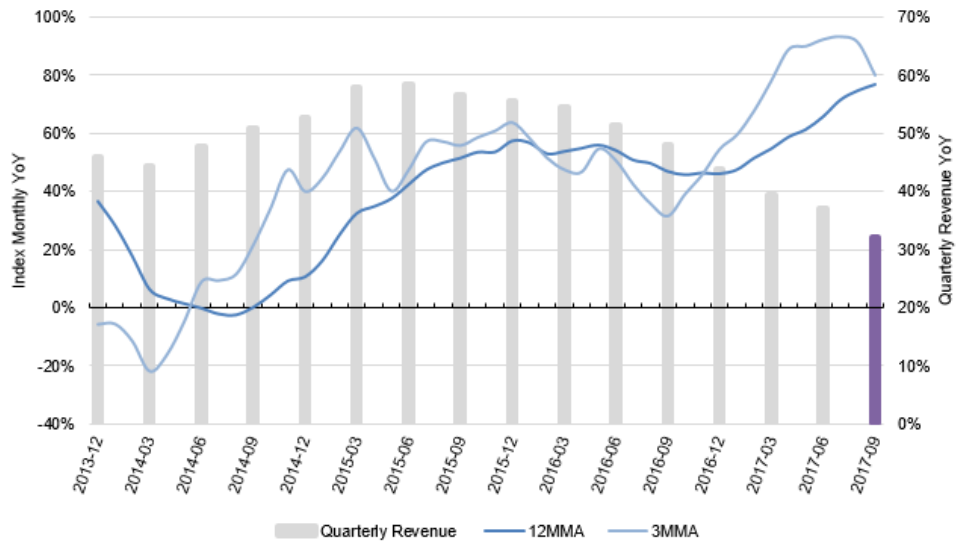
Eagle Alpha's Data Analytics team has published 36 quarterly reports for consumer companies incorporating Google search data. 25 of these indicators proved accurate equating to a hit rate of 69%.

Case Study

On 11th October 2017, Eagle Alpha published a Data Analytics report titled "HubSpot: Alternative Data Signals Possible Revenue Inflection". Figure 41 below shows 3 and 12 month moving averages (MMA) of company job listings data against quarterly revenue numbers.

The correlation between these two time series diverged in late 2016 and we concluded: "With job listings accelerating to the upside in 2017, management views on business prospects appear to have turned bullish. We view this positively for revenue growth in coming quarters."

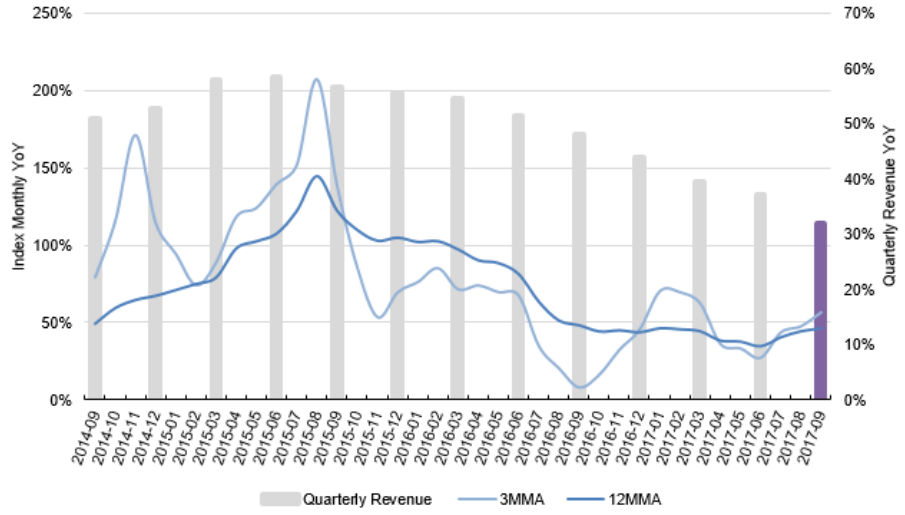
Figure 41: HubSpot Job Listings Index



Source: Eagle Alpha Analysis, Online Retail Data

Figure 42 below shows that our Search Signal indicator, built using HubSpot specific terms from Google Trends, experienced a similar trend as it bottomed in late 2016 and the 3 MMA crossed over the 12 MMA. We view this as a positive inflection point, but it can take several quarters to be reflected in company fundamentals.

Figure 42: HubSpot Search Signal



Source: Eagle Alpha Analysis, Online Retail Data

On November 3rd 2017, HubSpot reported better-than-expected Q3 results with revenue growing by 38% which was driven by 40% subscription revenue growth. The company also raised full year guidance ($370M to $371M vs consensus of $367.94M).

20. Equity (Expedia EXPE) > Discretionary (Long-Term) > Online Pricing

Key Takeaway

A web data provider accurately predicted that EXPE would miss 3Q17 room-night growth expectations, based on decelerating trends in reservation growth it started flagging in August. Its room-night estimate was within 1pp of reported.

Dataset

For EXPE, the provider analyzes comprehensive gross reservations, active properties, and bookable rooms on the Expedia.com and Hotels.com websites. Its systems are designed to visit the sites each day and find all active properties, then observe key facts for each property including location, property type, and booking messages, e.g. “10 people booked this property in the last 48 hours”. The data includes reservations from all EXPE brands (Expedia.com, Hotels.com, Travelocity, Orbitz, etc.) and is granular down to the property level.

Backtesting/Significance

The provider has been analyzing EXPE’s gross reservation data since mid-2015. Gross reservations refer to distinct accommodation transactions that are pre-cancellation and do not incorporate length of stay. Because gross reservations are captured at the time of booking, versus EXPE reporting at the time of stay, the data is a leading indicator of quarterly trends. The provider assumes a typical 1-month lag between time of book and time of stay.

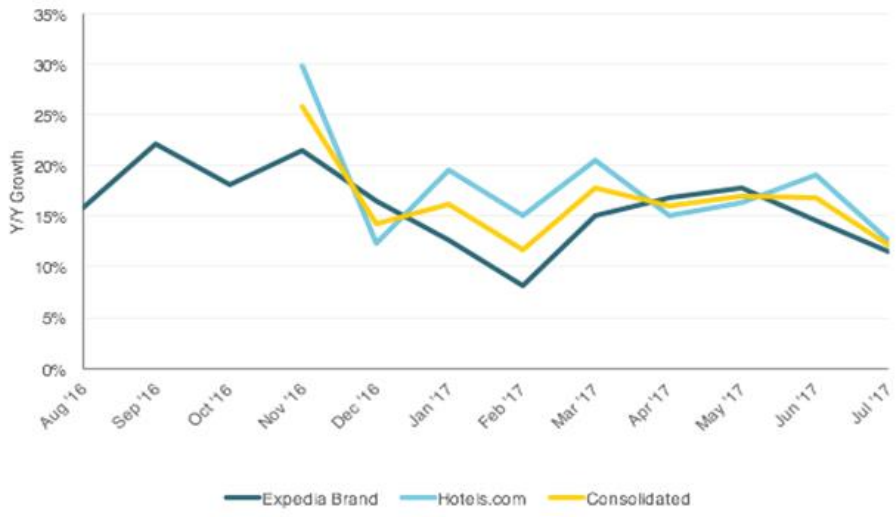
Gross reservation data has accurately predicted room-night growth within 2pp over the past three quarters.

Case Study

In August 2017, the provider reported that based on June and July data, EXPE’s core OTA business (i.e. ex-HomeAway) was seeing worsening trends in room-night growth. Both Hotels.com and what they refer to as Expedia brand (Expedia.com + regional brands) saw a slowdown in July, while Expedia brand had been deteriorating since May. See Figure 43.

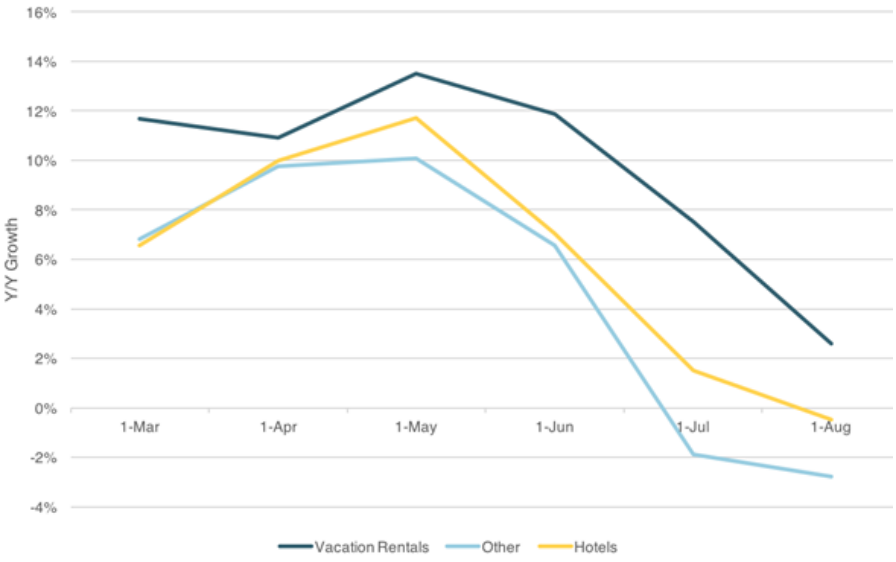
In its September update, the provider indicated that growth remained soft, with core OTA room-nights to come in at 13-15% Y/Y in 3Q17. The major driver of the pressure continued to be Expedia brands, as Hotels.com strengthened, with US destination reservations (the region represents roughly half of room-nights) dropping to flat Y/Y. See Figure 44.

Figure 43: Expedia Brand and Hotels.com Y/Y Growth in Gross Reservations Decelerated in July



Source: Online Pricing Data Provider

Figure 44: Expedia's Y/Y Reservation Growth Rate by Property Type, US



Source: Online Pricing Data Provider

The final 3Q17 preview indicated that core OTA trends remained light in September, suggesting a slow start to 4Q17 room-nights as well. The data provider also flagged in its preview that metasearch rebalancing efforts appeared to be a headwind in 3Q and that HomeAway would likely be the one bright spot due to continued monetization success. Both trends were consistent with company 3Q17 conference call commentary.

Disclaimer

This case study is directly from the vendor and Eagle Alpha has not backtested the data.

21. Equity (Square SQ) > Discretionary (Long-Term) > Consumer Transaction Data

Key Takeaway

Analysis of the email receipt data for Square indicated that the growth in number of sellers has been in decline since the first quarter of 2016.

Dataset

The dataset is delivered through a partnership with a provider that collects anonymized purchase data from around 2 million active shoppers, scanned from email purchase receipts. It covers over 600 merchants from more than 25 industries.

The provider transforms multi-form unstructured email receipt data into a normalized and digestible consumer transaction dataset. The dataset is also granular as it includes item and SKU-level transaction data, which is filtered into 53 product categories.

Geography	Coverage	Mapped to Tickers	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
USA	600 merchants	No	Since 2013	Real-Time	Weekly	7 days	Excel, TSV

Backtesting/Significance

Eagle Alpha's predictive model for Square has a mean absolute percentage error (MAPE) of 1.4%, compared to a consensus error of 4.7%. The standard deviation of our error is 1.6%.

Eagle Alpha's Data Analytics team has published 10 quarterly reports incorporating predictive indicators using the US Email Receipt Data. 8 of these indicators proved accurate equating to a hit rate of 80%.

Case Study

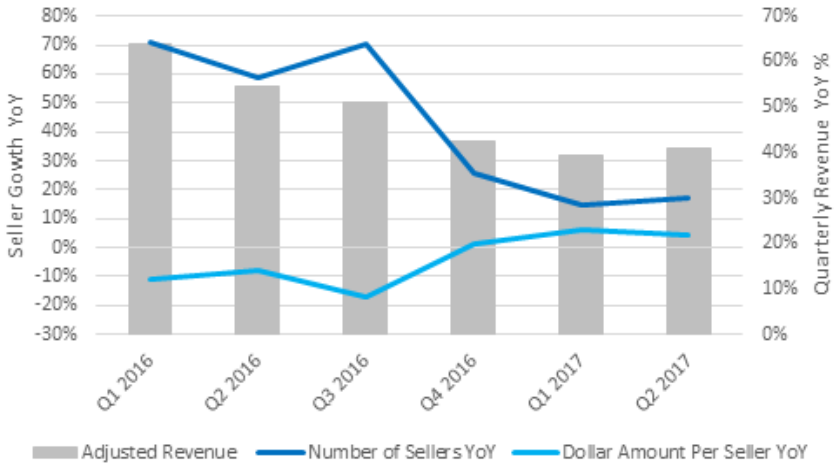
Square held an analyst day in May 2017 where the company gave details on its growth strategies. Two key parts for future growth are to move the company up market to get larger merchants, or sellers, to use Square devices and secondly to grow and retain existing merchants.

Larger sellers are users that have greater than $125k in annualized gross payment volume (GPV). In the June quarter of 2017, Square reported stronger than expected revenue. Adjusted revenue ticked up sequentially to 41% YoY growth, from 39% in the March quarter.

In our Data Analytics report, titled "Email Receipt Data Reveals Improving Seller Fundamentals", Eagle Alpha examined historical trends for Square sellers and dollar spend at sellers using an email receipt dataset. Square does not disclose seller metrics in annual or quarterly reports.

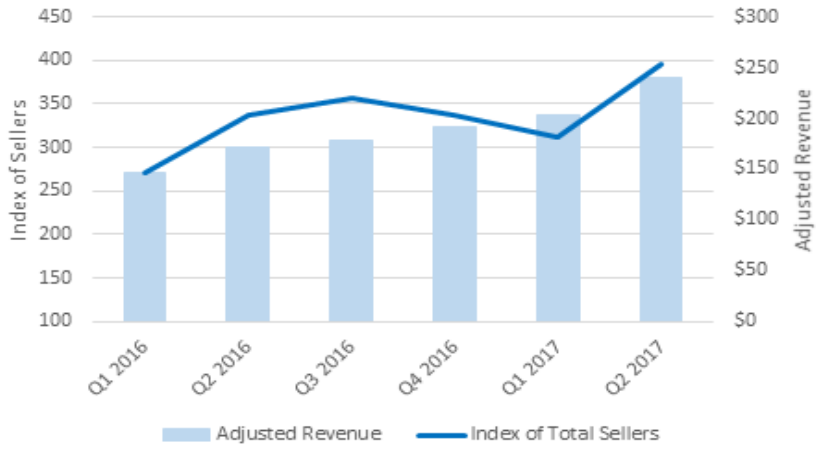
Analysis of the email receipt data indicated that the growth in number of sellers has been in decline since the first quarter of 2016. This has been the primary driver of slowing revenue growth for Square over this period (Figure 45). However, in the June quarter the number of sellers using Square's platform increased sequentially from 15% in March to 17%. This can also be seen in the unique seller index, where sellers increased to 396 from an average of 325 for all of 2016 (Figure 46).

Figure 45: Metrics for Sellers Improving



Source: Eagle Alpha Analysis

Figure 46: Unique Seller Index



Source: Eagle Alpha Analysis

From the email receipt data, we can also observe that average spend per seller was declining on a YoY basis during 2016. However, growth in spend per seller turned positive in Q4 2016 and this improvement has been helping revenue growth over the first half of 2017.

Data on growth in total sellers, and the amount sold per seller, may be an indication that growth strategies outlined by the company are starting to gain traction and are a positive sign for longer term fundamentals of the company.

On November 8th 2017, the company reported strong third quarter results and raised guidance for the year. Full year adjusted revenue was increased to $963m-$966m from $925m-$935m.

22. Equity (Finish Line FINL) > Discretionary (Long-Term) > Online Search Data

Key Takeaway

The search indicator correctly predicted weakness in SSS ahead of earnings in December 2016. The company's stock reacted to the change in fundamentals and is down significantly since then.

Dataset

Google Trends is a public web tool based on Google Search that shows how often a particular search term is entered relative to the total search-volume over time across various regions of the world. Using Google Trends, Eagle Alpha has built company specific indices based on search terms that are related to a given retailer's product offering. This involves an exhaustive process for identifying search terms related to a company's revenues using both internal and third party tools.

Geography	Coverage	Mapped to Tickers	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
Global	B2B, Retail, Luxury, Restaurants	Yes	Since 2006	Daily, Weekly and Monthly	Monthly	1-3 days	JPG, PNG, SVG, PDF, CSV

Backtesting/Significance

Observing crossing points of three-month and one-month moving averages of one of these search indices has proved predictive of inflection points in revenue growth for companies across a broad range of sectors including retail, luxury goods and restaurants.

Over the last 18 quarters the Finish Line indicator demonstrated a three-quarter hit rate of 78% in in-sample testing. This means that 78% of the time the three-month moving average has moved in the same direction as same store sales over a three-quarter period.

Note: Eagle Alpha's Data Analytics team has published 36 quarterly reports for consumer companies incorporating Google search data. 25 of these indicators proved accurate equating to a hit rate of 69%.

Case Study

On December 5th 2016, Eagle Alpha published a report on US sports retailer Finish Line using Google Trends data. The Figure 47 below shows the signal pointing to a strong uptrend prior to the August 2016 quarter when the company reported better than expected results.

However, the directional change of the index in the November 2016 quarter indicated that Finish Line's SSS (same-store sales) growth was at risk and Eagle Alpha data insight analysts anticipated that management outlook for the following quarter could disappoint.

On December 21st 2016, Finish Line reported a weak quarter and gave poor guidance for the following quarter. "We are disappointed that our third quarter sales and earnings fell short of our expectations," said Sam Sato, Chief Executive Officer of Finish Line.

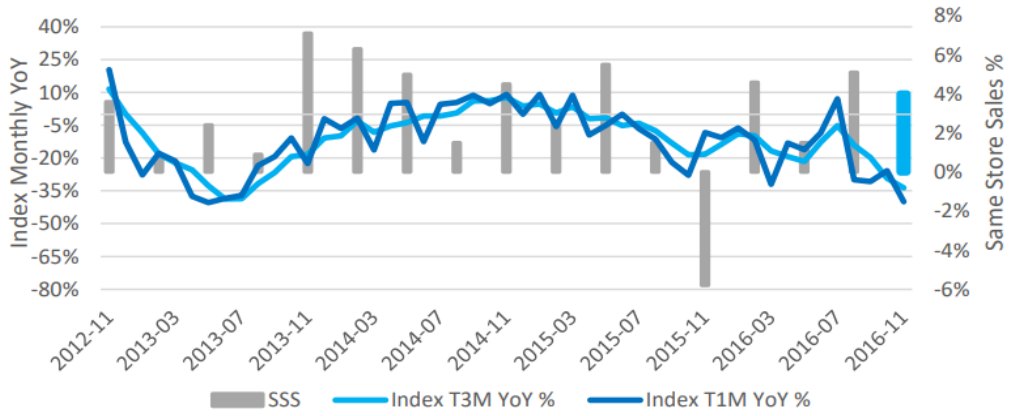
Eagle Alpha's Data Analytics report on December 5th 2016, indicated that the search index, an indicator of consumer demand based on online search data, was pointing to weakness in SSS for Finish Line. The report stated: "the sharp downtrend observed in the index in recent months (Figure

47) indicates that expectations for SSS growth at FINL are at risk, or that management outlook for the February quarter may disappoint".

Finish line reported SSS of 0.7% for the November 2016 quarter compared to expectations for growth in excess of 4%. The company also reported a significant drop in SSS in the February quarter of -4.5%.

The search indicator correctly predicted weakness in SSS ahead of earnings. The company's stock reacted to the change in fundamentals and is down significantly since then.

Figure 47: Search Signal Index for FINL



Source: Eagle Alpha Analysis

23. Equity (Burberry BRBY) > Discretionary (Long-Term) > Online Search Data

Key Takeaway

Citi concluded that the short-term 1-month YoY observation crossing over the 3-month moving average YoY indicates major inflection points of same store sales growth for Burberry.

Dataset

Google Trends is a public web tool based on Google Search that shows how often a particular search term is entered relative to the total search-volume over time across various regions of the world. Using Google Trends, Eagle Alpha has built company specific indices based on search terms that are related to a given retailer's product offering. This involves an exhaustive process for identifying search terms related to a company's revenues using both internal and third party tools.

Geography	Coverage	Mapped to Tickers	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
Global	B2B, Retail, Luxury, Restaurants	Yes	Since 2006	Daily, Weekly and Monthly	Monthly	1-3 days	JPG, PNG, SVG, PDF, CSV

Backtesting/Significance

Observing crossing points of three-month and one-month moving averages of one of these search indices has proved predictive of inflection points in revenue growth for companies across a broad range of sectors including retail, luxury goods, restaurants and 2B software.

Over the last 19 quarters the Burberry indicator demonstrated a two-quarter hit rate of 58% in in-sample testing, and three-quarter hit rate of 68%. This means that 58% of the time the three-month moving average has moved in the same direction as same store sales over a two-quarter period.

We prefer to measure the accuracy of our search signals indicators by reference to the hit rate, as we believe it better captures the goal of the tool to identify inflection points in growth for a company. However, below we present analysis from Citi's March 2017 report "Searching for Alpha: Big Data" which focuses on correlations.

Figure 48 appears to support our prior that short-term 1-month YoY observation crossing over the 3-month moving average YoY indicates major inflection points of same store sales growth as depicted by the green circles on the chart. Another interesting finding is that consensus[25] is pretty bad at predicting same store sales growth. In fact, it only achieves close to 20% of correlation with actual reported figures. With either the 1-month YoY or 3-month moving average YoY measures based on EA equity index for Burberry, the correlation jumps to over 70% which is a significant improvement. The additional advantage of this search data is its timeliness – the data at the end of the quarter is available immediately, whilst official figures typically are announced at least 3 weeks after quarter-ends. The timeliness and much improved correlation of the new dataset with actual reported figures make such an offering appealing.

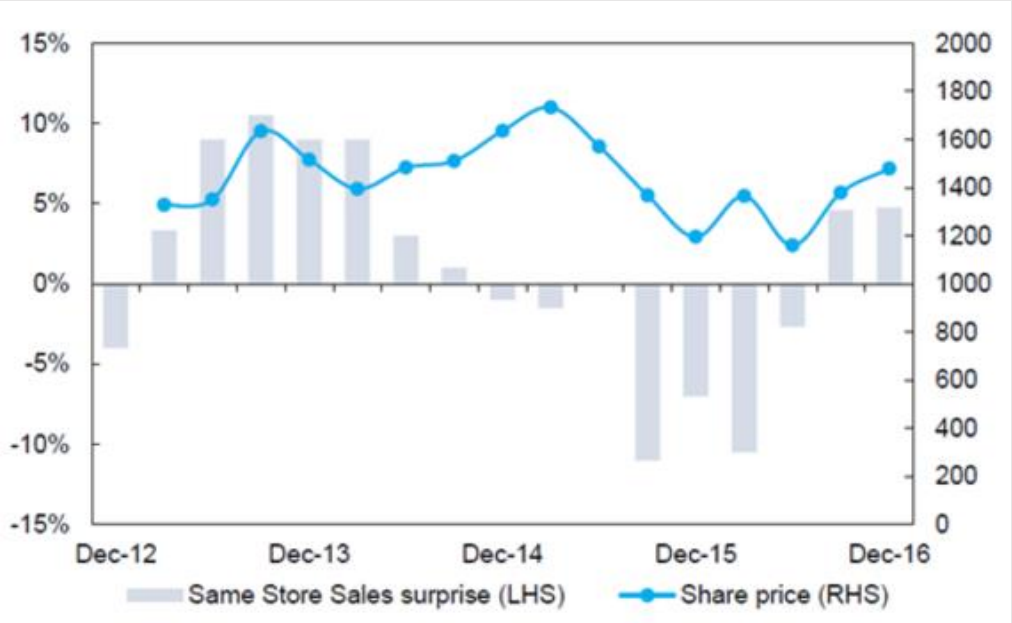
[25] *The consensus data is sourced from Bloomberg but it has low analyst coverage issues as not all analysts publish their estimates for same store sales.*

In Figure 49 we highlight that same store sales does have pricing impact especially the surprise element of it. That is, markets react to the positive/negative sales surprise. This suggests that, if we are able to predict the inflection points better with Google Trends data, there could be pricing implications from being able to act sooner and more accurately than the bulk of investors.

Figure 48: Burberry Same Store Sales vs Eagle Alpha Stock Index

Source: Eagle Alpha, Citi Research

Figure 49: Burberry's Same Store Sales Surprise vs Share Price



Source: Bloomberg, Citi Research

24. Equity (Sportswear Industry) > Discretionary (Long-Term) > Online Search Data

Key Takeaway

Online search data provided early indicator of weakness in sportswear sector. The industry analysis supported our case for fundamental weakness for FINL and FL and points to longer term fundamental issues for the sportswear space.

Dataset

Google Trends is a public web tool based on Google Search that shows how often a particular search term is entered relative to the total search-volume over time across various regions of the world. Using Google Trends, Eagle Alpha has built company specific indices based on search terms that are related to a given retailer's product offering. This involves an exhaustive process for identifying search terms related to a company's revenues using both internal and third party tools.

Geography	Coverage	Mapped to Tickers	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
Global	B2B, Retail, Luxury, Restaurants	Yes	Since 2006	Daily, Weekly and Monthly	Monthly	1-3 days	JPG, PNG, SVG, PDF, CSV

Backtesting/Significance

Observing crossing points of three-month and one-month moving averages of one of these search indices has proved predictive of inflection points in revenue growth metrics for companies across a broad range of sectors including retail, luxury goods and restaurants. Over a nineteen quarter period a Search Signal for Finish Line (FINL) and Foot Locker (FL) has demonstrated a three-quarter hit rate of 74% in in-sample testing. This means that 74% of the time the three-month moving average has moved in the same direction as same store sales (SSS) over a three-quarter period. Over a twelve quarter period the indicator for Under Armour (UAA) has demonstrated a three-quarter hit rate of 67% in in-sample testing.

Note: we have published 36 quarterly reports for consumer companies incorporating Google search data. 25 of these indicators proved accurate (69% hit rate).

Case Study

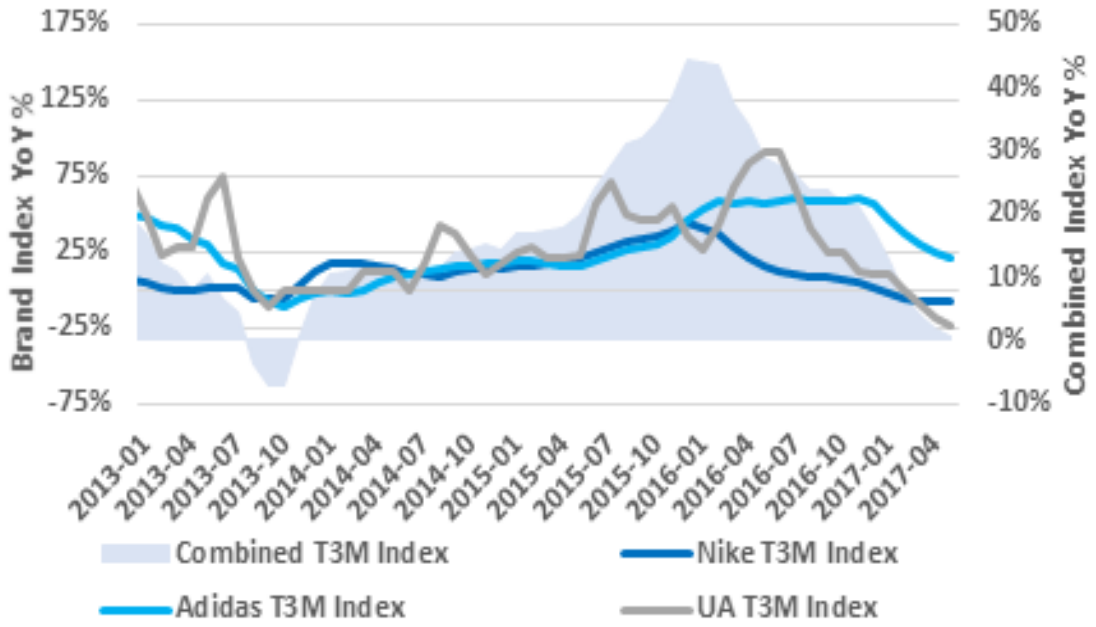
In December 2016, Eagle Alpha published a report on US sports retailer FINL and FL using Google Trends data. The Search Signal for Finish Line was indicating that SSS growth was at risk and could disappoint management expectations. Subsequently Finish Line reported a poor quarter and offered weak guidance for the March quarter. In February of 2017 Eagle Alpha published a follow up report on the sports retailers where we highlighted that the Search Signal for FL had deteriorated and pointed to a risk of the company missing expectations. The company reported strong SSS and offered positive guidance for the April quarter. However, in April the company preannounced negatively for the quarter, reducing SSS guidance from 5% to 2%. When the company reported the quarter SSS was just 0.5%. Search Signals proved to be an early indicator of fundamental weakness for both FL and FINL.

The negative trends seen at FL and FINL triggered Eagle Alpha to take a deeper dive into the industry to see what could be catching company management off guard. Analysis of industry trends led to the

conclusion that the problems for FINL and FL were more rooted in industry dynamics than company specific issues. This can be seen in Figure 50. Using Google Trends, we created indices of sports apparel and sports footwear for the big three global sports brands of Adidas, Nike and Under Armour.

We also created a combined index of all three brands. As can be seen in Figure 50, a three-month moving average of our index began to roll over in early 2016, stabilized somewhat in mid-2016 and then accelerated to the downside in late 2016. The decline in the combined index happened in waves. This can be observed by looking at the individual brands. Nike rolled over in early 2016, followed by Under Armour in mid-2016 and finally by Adidas in late 2016.

Figure 50: Big 3 Global Brands in Decline Through 2016



Source: Eagle Alpha Analysis

This industry analysis supported our case for fundamental weakness for FINL and FL and pointed to longer term fundamental issues for both companies. The Search Signal indicators for both companies remained weak in June, indicating further potential weakness in SSS. Subsequently both FL and FINL reported very weak quarters and the stocks traded off significantly. From the time of Eagle Alpha's first Search Signal report on FINL and FL in December of 2016 to September of 2017 the stocks declined 65% and 55% respectively.

Investors have only recently started to consider the implications for Adidas and Nike of the negative results at the retailers and Under Armour. These stocks have been weak in recent months and have seen broker downgrades. However, headwinds for the sector were apparent in search data over 12 months before.

25. Equity (Time Warner TWX) > Discretionary (Long-Term) > Mobile App Data

Key Takeaway

App data showed an early indicator of a positive inflection in revenue growth for HBO, one of Time Warner's largest divisions.

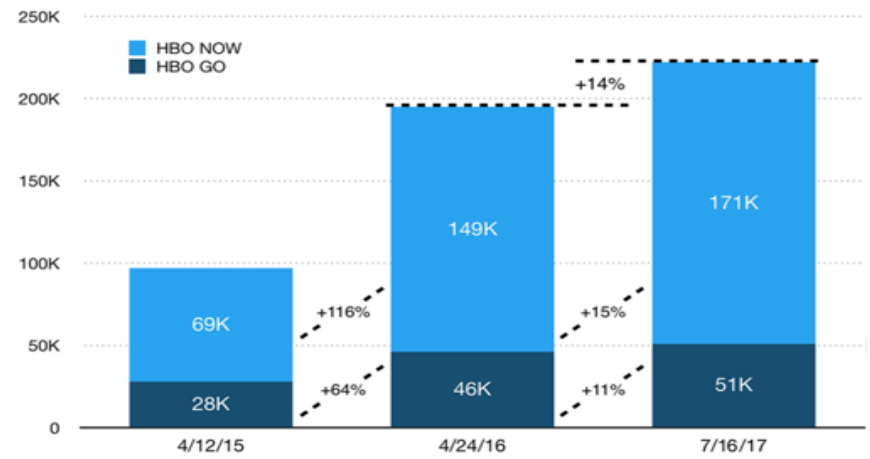
Dataset

App download and usage data is an example of an alternative data source that can be used to analyze product adoption and brand satisfaction. This type of data enables investors to track activity in a wide range of sectors from banking to food delivery to online entertainment.

Case Study

On July 19th 2017, an app data provider published a report highlighting record number of new installations of both HBO NOW and HBO GO. The network's streaming services reached the peak of approximately 222,000 installs due to the premiere of season seven of Game of Thrones (Figure 51).

Figure 51: HBO Mobile App Peak New Installs (United States)



Source: App Data Provider

The subsequent report publish on August 18th stated: "The sizable influx of new subscribers reflected in our download estimates at the time drove record single-day revenue one month later as their 30-day free trials converted to paid subscriptions." HBO NOW rose to the top of the App Store revenue charts and grossed around $2.6 million across both the App Store and Google Play on August 16th 2017. This represents a 41% increase over the previous record of $1.3 million recorded on June 24th 2016.

HBO is one of Time Warner's largest divisions but growth for the division has been slowing in recent quarters, therefore this analysis could be an early indicator of a positive inflection in revenue growth for HBO.

Disclaimer

This case study is directly from the vendor and Eagle Alpha has not backtested the data.

26. Equity (Activision Blizzard) > Discretionary (Long-Term) > Social Media Data

Key Takeaway

Using social media data, we correctly highlighted that the Overwatch game was well positioned to set a new sales record for Activision Blizzard (ATVI).

Backtesting/Significance

Eagle Alpha's Data Analytics team has published 15 quarterly reports incorporating social data. 11 of these indicators proved accurate equating to a hit rate of 73%.

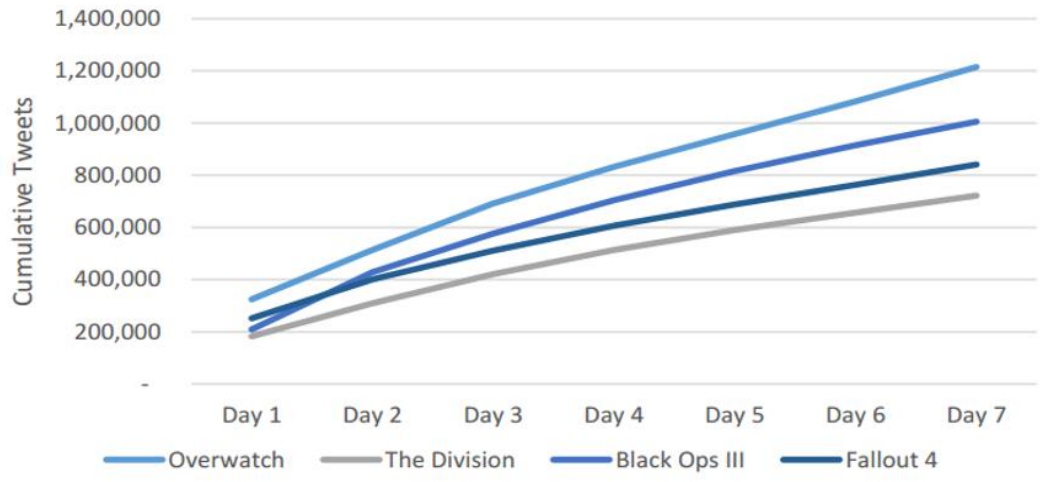
Dataset

Social media data is created through the public's interaction with social media platforms such as Twitter, Facebook, YouTube and Twitch. Access to this data can be obtained directly from the provider themselves or often through third party platforms. This case study also incorporated Eagle Alpha's Web Queries tool. Web Queries is a query based tool that enables clients to obtain analytics on over 90 million web sources including: blogs, image and video sites, forums, review sites, social media and news sites.

Case Study

On June 3rd 2016, Eagle Alpha published a Data Analytics report on video games publisher Activision Blizzard (ATVI US) titled "Activision Blizzard: Overwatch Positioned to Set New Sales Record." The report highlighted the potential success of the company's latest game, Overwatch. Overwatch enjoyed stronger Twitter visibility when compared to the competition and the strong positive sentiment towards the title pointed to positive consumer reaction to the title.

Figure 52: Overwatch Twitter Visibility Much Stronger than Comp Titles

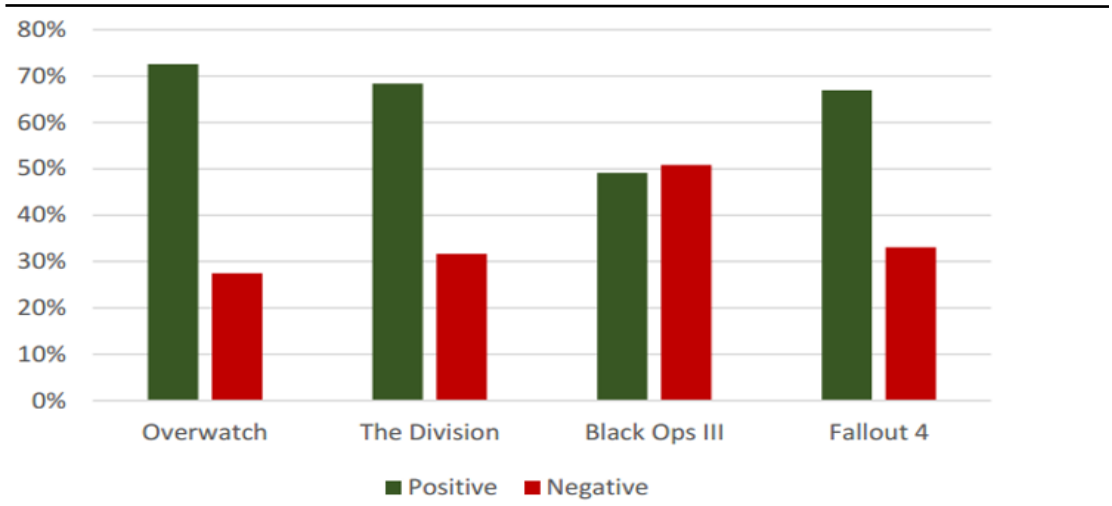


Source: Eagle Alpha Web Queries

Over 1.2m tweets mentioned Overwatch in the first week post the game's launch, while the competition gathered much lower number of mentions: Black Ops III (1m), Fallout (840k) and the Division (720k).

The Eagle Alpha Web Queries tool was also used to analyse consumer comments and sentiment around Overwatch. The first week comparative statistics are presented below, and clearly show a stronger performance for Overwatch compared to rival titles.

Figure 53: Overwatch Leads on Twitter Sentiment for First Week



Source: Eagle Alpha Web Queries, Twitter

On June 14th 2016, Activision reported that Overwatch had over 10 million players after only three weeks of the launch. The Division, on the other hand, was launched in the beginning of March 2016 and had 9.5 million players more than two months after the launch.

In August 2016, Overwatch hit the 15 million user mark, reaching the milestone faster than any other game in Blizzard's history. More recently, January 27th 2017, Activision announced that Overwatch reached another milestone by amassing more than 25 million players.

27. Equity > Discretionary (Long-Term) > Trade Data

Key Takeaway

The South Korea real-time export data accurately tracked revenue of construction machinery companies throughout the Q1'12 – Q2'17 period.

Dataset

This vendor delivers high frequency and comprehensive South Korea export data. Preliminary export data includes total volume and value of all products, at all country destinations that are exported from South Korea. The dataset is created by aggregating and analyzing customs declaration forms. History: Since 2003.

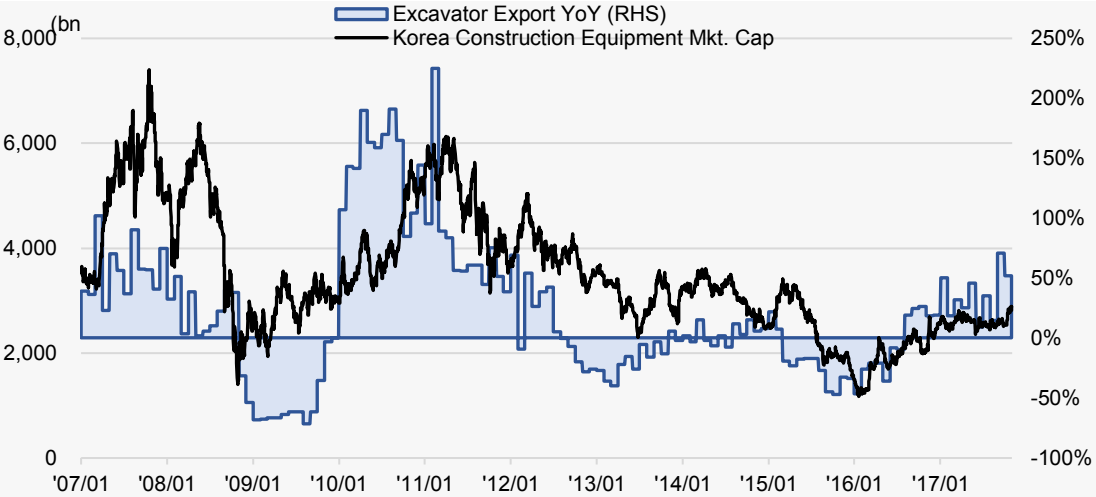
Backtesting/Significance

The South Korea real-time export data shows R-squared of 0.86 with revenues for construction machinery companies over a 5 year period and R-squared of 0.75 with YoY revenue growth. The calculated export value also correctly projected the directional movement of reported revenues for construction machinery companies in all quarters between Q1 2012 and Q2 2017.

Case Study

Monthly export value of excavator showed correlation of 82% with total market capital of S. Korean construction machinery companies from Jan. 2012 ~ Sep. 2017. Data was also indicative to Volvo's (STO:VOLV-A) market cap. with 62% correlation to S. Korean excavator export value. Volvo's Changwon plant accounts 40% of group's total revenue and exports key components to 8 other production facilities.

Figure 54: Excavator Qtr Export vs Revenue Aggregate of Related Companies



Source: Trade Data Provider

Disclaimer

This case study is directly from the vendor and Eagle Alpha has not backtested the data.

28. Equity (Lululemon LULU) > Discretionary (Long-Term) > Online Search, Social Media, Pricing Data

Key Takeaway

Eagle Alpha's analysis proved correct, i.e. Lululemon (LULU) reported sales growth of 13% YoY in Q3 2016 which was in line with our expectations.

Datasets

Google Trends is a public web tool based on Google Search that shows how often a particular search-term is entered relative to the total search-volume over time across various regions of the world.

Social media data is created through the public's interaction with social media platforms such as Twitter, Facebook and YouTube. Access to this data can be obtained directly from the provider themselves or often through third party platforms.

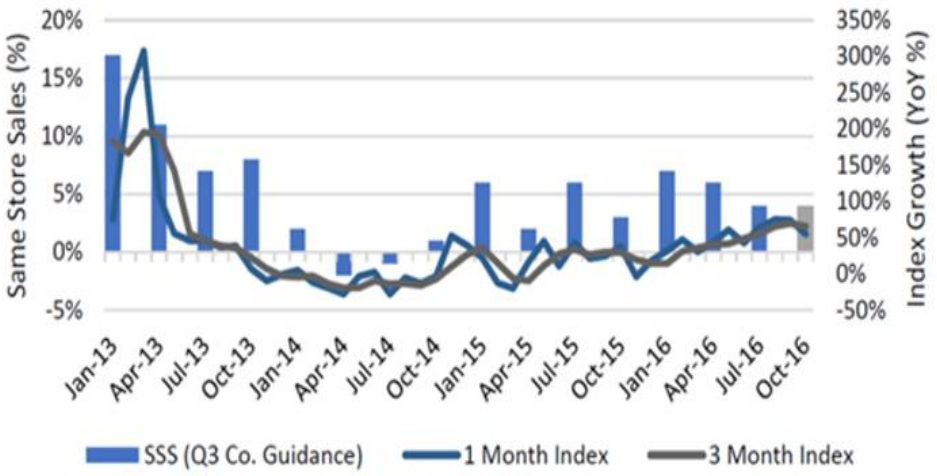
Pricing data is crawled from the websites of large retailers. This can provide an insight into long term trends, as well as the most recent trading performance, and is based on metrics such as average selling price and share of bestselling products in a category. Eagle Alpha owns this dataset.

This case study also incorporated Eagle Alpha's Web Queries tool. Web Queries is a query based tool that enables clients to search over 90 million web sources including: blogs, image and video sites, forums, review sites, social media and news sites.

Case Study

Lululemon's (LULU) same-store sales (SSS) have been decelerating since the beginning of 2016 with Q2 numbers particularly disappointing. On November 22nd 2016, Eagle Alpha published a note on Lululemon highlighting stabilized momentum in Q3. Figure 55 below shows that search data suggested stable to improving SSS in Q3 2016.

Figure 55: Search Signal Index for LULU

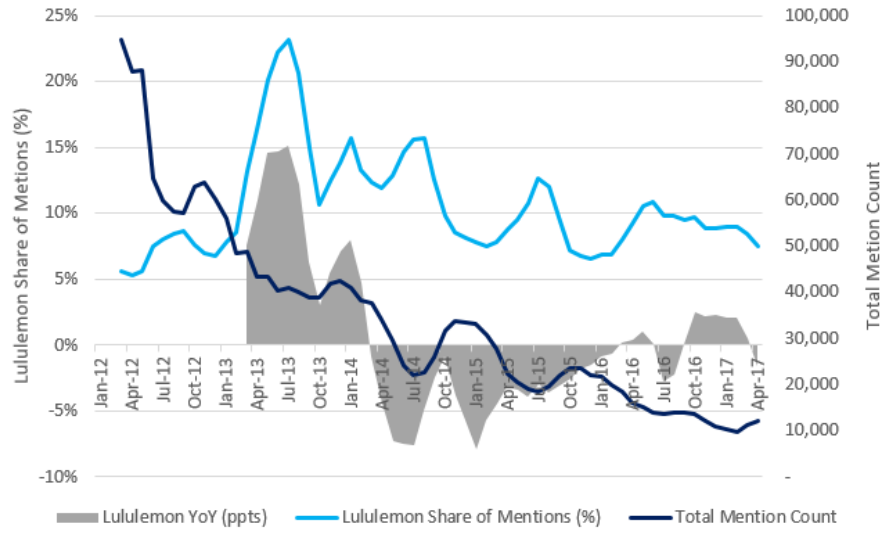


Source: Eagle Alpha Analysis

According to social media data, the athleisure trend was experiencing continued pressure. However, Lululemon improved its competitive position as the company received a meaningful share of mentions across blogs and forums for the first time since 2014 (Figure 56).

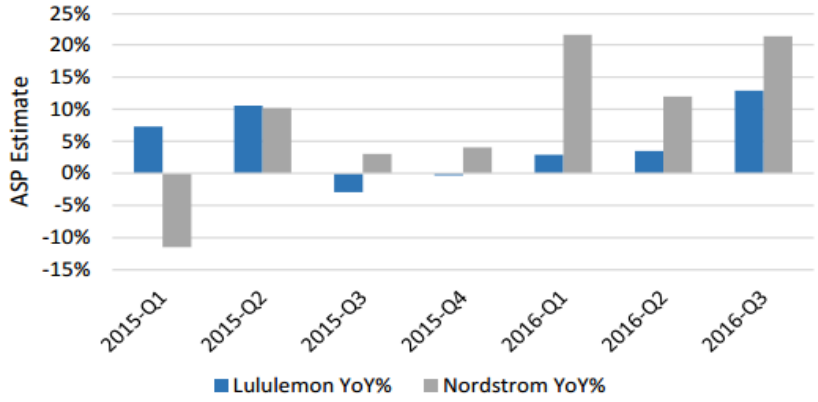
Eagle Alpha's YoY average selling price (ASP) estimate for Lululemon grew by 13% in Q3 2016, representing the fastest change since Q1 2015 (Figure 57).

Figure 56: Lululemon Share of Mentions Across Blogs and Forums



Source: Eagle Alpha Web Queries

Figure 57: Lululemon Price Growth Accelerated



Source: Eagle Alpha Analysis, Online Pricing Data

In our November 2016 report, we concluded: "consensus estimates are for Q3 2016 revenues growth of 13% YoY, slightly behind the 14% number reported in Q2. This revenue growth looks achievable." On December 7th 2016, Lululemon reported sales growth of 13% YoY – in line with our expectations.

29. Equity > Discretionary (Long-Term) > Online Reviews

Key Takeaway

Companies that get better reviews from employees post better share price performance.

Dataset

This vendor operates a website where employees and former employees review companies and management of companies on an anonymous basis.

Case Study

The vendor researched the performance of 'Best Places to Work' companies. Companies getting the award have satisfied employees and also turned out to be good investments. The table below shows the relative performance of an equally-weighted portfolio of the 26 public companies included in the 2016 Best Places to Work list,

Figure 58: Stock Returns for 2016 Best Places to Work Companies vs. the S&P 500

Employer Name	January 2, 2015 Stock Price	November 18, 2015 Stock Price	Percentage Gain (Loss)
Paycom	$26.06	$42.00	61.2%
Hubspot	$33.48	$52.91	58.0%
Expedia	$85.76	$125.30	46.1%
Google	$529.55	$760.01	43.5%
Facebook	$78.45	$107.77	37.4%
NIKE	$95.03	$125.78	32.4%
Salesforce	$59.24	$77.35	30.6%
Adobe	$72.34	$91.27	26.2%
Guidewire	$49.97	$58.62	17.3%
Red Hat	$68.99	$80.53	16.7%
Costco Wholesale	$141.61	$161.25	13.9%
Southwest Airlines	$42.69	$46.31	8.5%
LinkedIn	$229.65	$248.76	8.3%
Apple	$109.33	$117.29	7.3%
Gartner	$83.60	$87.87	5.1%
Workday	$80.41	$83.68	4.1%
Stryker	$93.99	$97.01	3.2%
Delta Air Lines	$49.18	$47.75	-2.9%
Akamai	$63.25	$59.25	-6.3%
Eastman Chemical	$76.48	$71.62	-6.4%
Vivint Solar	$9.24	$7.99	-13.5%
Zillow	$32.40	$26.79	-17.3%
Chevron	$112.58	$92.21	-18.1%
F5 Networks	$130.33	$104.13	-20.1%
Twitter	$36.56	$25.90	-29.2%
SolarCity	$52.92	$27.75	-47.6%
2016 Best Places to Work Portfolio Return			***9.9%***
S&P 500 Return			***1.2%***

Source: ESG Data Vendor

30. Equity > Discretionary (Long-Term) > Online Reviews

Key Takeaway

The assessment of service and product quality of auto lenders has proven to be indicative of deeper operational issues which are then reflected in stock prices.

Datasets

There are three datasets of relevance:

1. Consumer Affairs: The dataset contains consumer complaints or reviews and ratings for over 3,300 U.S companies and product lines. History: since January 1999.

2. Better Business Bureau: The BBB serves as an intermediary between consumers and businesses in consumer disputes, as well as collecting and providing business ratings. History: since January 2012.

3. Consumer Finance Protection Bureau: This U.S. government regulatory agency provides access to its database of customer complaints relating to financial services companies. History: since January 2011.

Case Study

On October 13th 2015, Eagle Alpha's Data Analytics team published a research note on U.S. auto lenders highlighting customer service issues. Santander Consumer (SC) had the worst consumer complaint and review profile, followed by Ally Financial (ALLY). The two companies had:

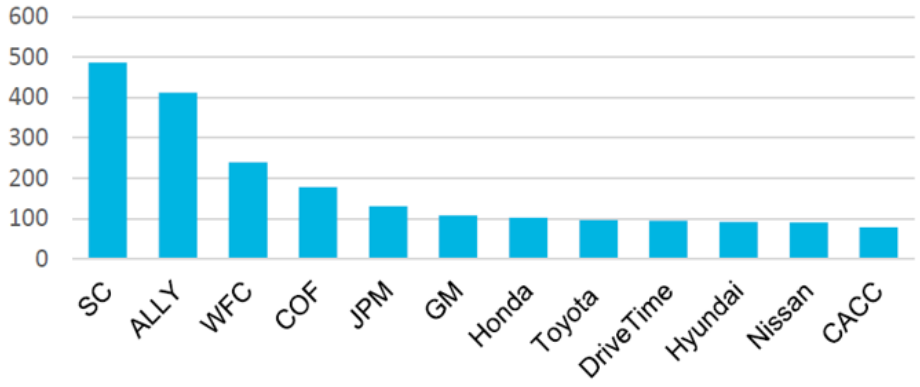
• The highest number of complaints to Consumer Finance Protection Bureau (CFPB).

• The highest number of complaints to Better Business Bureau (BBB).

• The highest number of poor reviews on Consumer Affairs website.

• Among the lowest ratings from consumer finance websites.

Figure 59: Complaints to CFPB Regarding Vehicle Loans & Leases, 2015 YTD



Source: CFPB, Eagle Alpha. These figures were high in absolute terms and in Santander Consumer's case, significantly higher relative to its loan issuance.

Santander Consumer's stock price was down 45% nine months after the publication of our report. Part of this decline is attributable to a general selloff in auto lenders as the market is pricing in a cyclical peak in US auto sales. However, Santander Consumer has declined more than the other auto lending stocks in the 9 months after Eagle Alpha issued its report (Figure 60).

Although Santander Consumer's business suffers from a number of troubles not directly related to its poor reviews, including capitalization issues, the reviews and complaints were a "canary in the coalmine" signal, supporting the idea that there is value in understanding a company's quality ratings as it can affect operations and earnings.

Figure 60: Performance of SC Share Relative to the Peer Group since October 13, 2015

Source: Bloomberg

The assessment of service and product quality of auto lenders, in this case, has proven to be indicative of deeper operational issues which are then reflected in stock prices.

31. Macro > Discretionary (short-term) > Satellite Imagery Data

Key Takeaway

A relationship was identified between copper prices and estimates of copper inventories using satellite imagery data.

Dataset

Satellite data company provides investors with actionable insights on year-over-year traffic trends based on proprietary quantitative analysis of satellite data, retail parking lot imagery, store locations, and climate data.

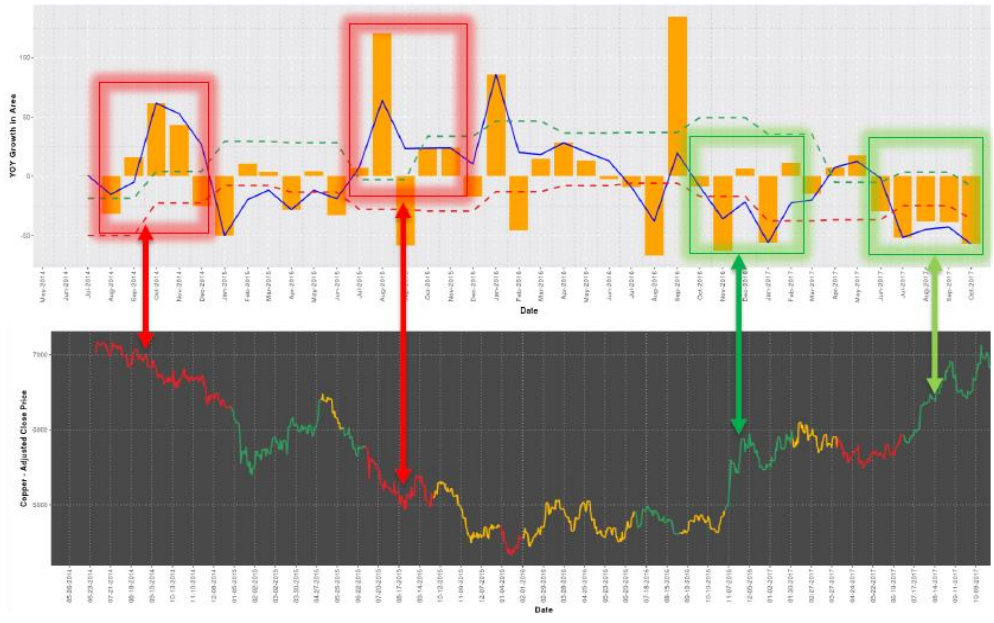
The traffic data can be used to compare the current y/y traffic growth data against the historical average of the y/y traffic growth measurements. A signal is generated when the current y/y traffic growth rate is above or below certain standard deviation levels, indicating a significant change in shopper behavior.

Case Study

During periods with significant increases in inventories (highlighted by red boxes in the figure below), the expectation is that copper prices are going to decrease.

During periods with significant decreases in inventories (highlighted by green boxes), on the other hand, the opposite is true with the expectation for copper prices to rise.

Figure 61: Satellite Imagery Metal Signals



Source: Satellite Data

Disclaimer

This case study is directly from the vendor and Eagle Alpha has not backtested the data.

32. Macro > Discretionary (Long-Term) > Online Search Data

Key Takeaway

Eagle Alpha's US Unemployment index has a 5-year correlation of 0.9 with the US Unemployment Rate, with an out-of-sample prediction improvement of 14% compared with a baseline ARIMA model.

Dataset

Google Trends is a public web tool based on Google Search that shows how often a particular search-term is entered relative to the total search-volume over time across various regions of the world. There are several advantages to using search data as a determinant of economic activity. The data is very timely, it has over 10 years of history, it is unique and offers unparalleled flexibility in terms of the variety of issues that can be analysed. The data is also generated as a by-product of people's normal day-to-day activity, as opposed to traditional survey methods which rely on individuals or firms responding to survey questions after the event. This can avoid problems associated with non-response or inaccurate responses.

Geography	Coverage	Mapped to Index	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
U.S	US Labor Market	Yes	Since 2006	Monthly	Monthly	1 to 14 days	CSV

Backtesting/Significance

Eagle Alpha data scientists and data insight analysts have invested three years into finding the best way to use online search data to predict economic indicators. We have devised our own proprietary methodology that leverages all relevant academic research, as well as accepted best practices in the field.

Each index is built using a rigorous process: 1) generate relevant search terms; 2) source the search volume for each term dating back to 2004; 3) clean the data and adjust for outliers and seasonality; 4) search terms are ranked by their predictive scores; and 5) final index includes a selected basket of terms, and measures co-movement of search activity with a particular economic indicator. To obtain a detailed overview of our methodology please contact us.

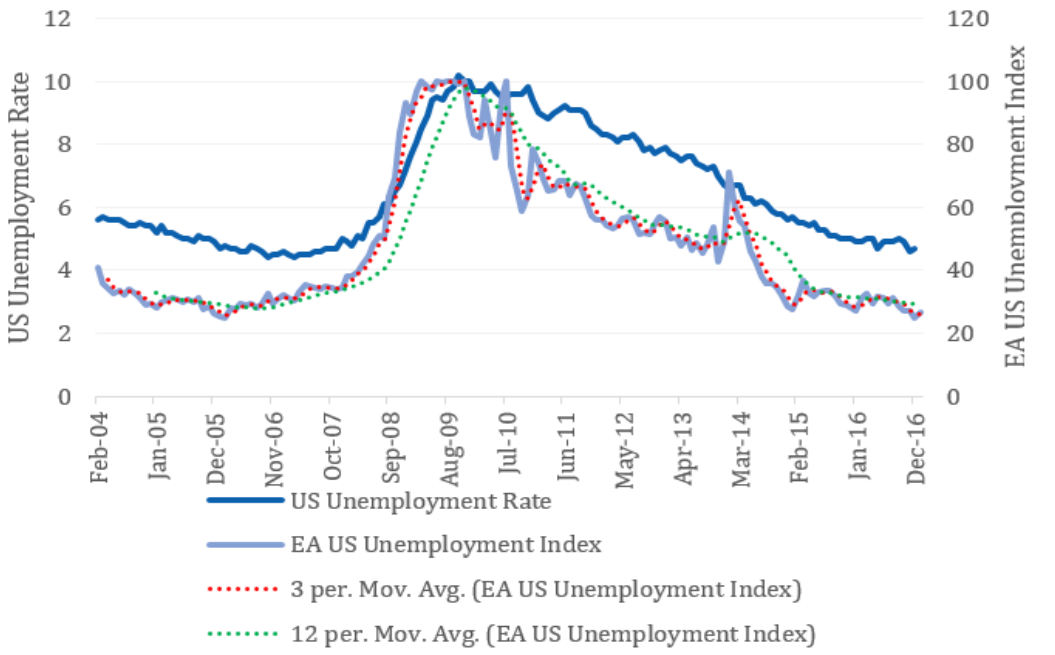
These indices are not designed to provide point estimates for macro investors, but instead add value by improving the predictive power of clients' estimation models.

Case Study

Eagle Alpha's US Unemployment Index (Figure 62) is a measure of online search activity relating to the claiming of unemployment benefits. The index has a 5-year correlation of 0.9 with the US Unemployment Rate and testing shows an out-of-sample prediction improvement of 14% over a baseline ARIMA model of unemployment over the same period.

In January 2017, the index increased by 1.8 points. This is the largest MoM change since August 2016. The 3 and 12-month moving averages mirror each other closely, which is consistent with the US economy operating close to full employment.

Figure 62: Eagle Alpha US Unemployment Index vs. Unemployment Rate (US)



Sources: Eagle Alpha, Google, BLS, Bloomberg

33. Macro > Discretionary (long-term) > Employment Data

Key Takeaway

Employment data enables more granular analysis of the labor market by sector.

Datasets

Completely unique in the industry, the job listing dataset only indexes jobs directly from employer websites. Updated daily with over 4 million jobs from more than 30,000 employers, the platform eliminates duplicate and expired job listings, as well as job pollution. From the core platform, the company has developed an array of products and services and achieved significant market traction in two primary business units: Candidate Sourcing and Job Market Data and Analytics.

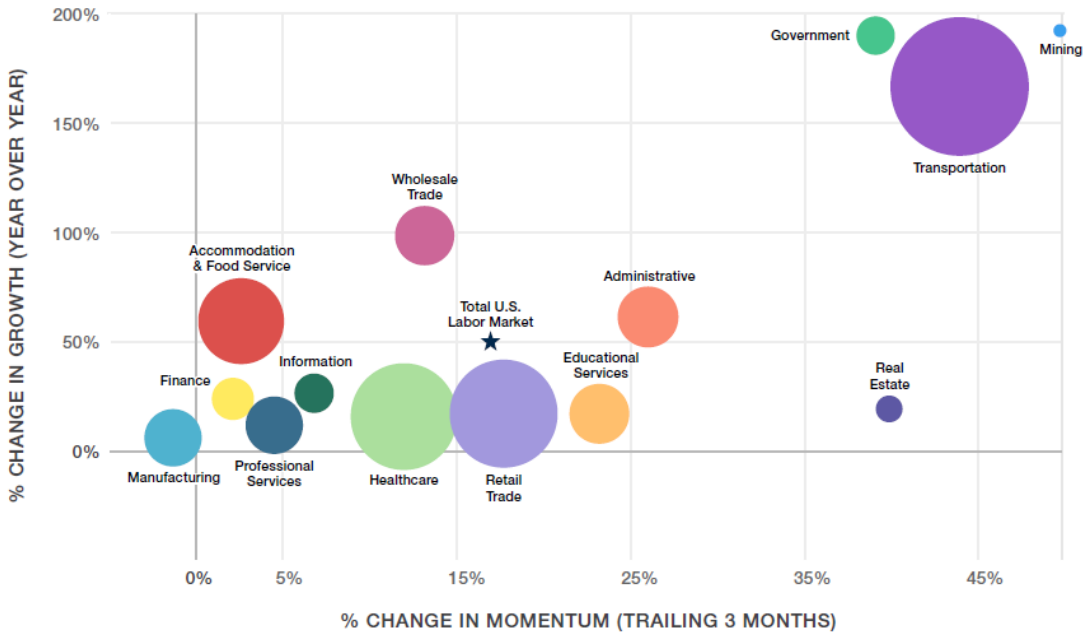
Backtesting/Significance

Backtesting of the dataset demonstrated alpha in the dataset, with "Jobs Active" producing the highest and most consistent returns. Yearly hedge returns were between 6-8%.

Case Study

Official figures by the end of August 2017 were showing low levels of unemployment in the US. However, employment data showed that not all industries were on the hiring spree. The figure below shows a sector-level look at the labor market, focusing on 12-month and three-month growth rates. Retail and manufacturing were bringing up the rear, while transportation, national security and mining were gaining momentum.

Figure 63: Change in Labor Demand by Industry



Source: Employment Data

34. Macro > Discretionary (Long-Term) > Trade Nowcasting Data

Key Takeaway

Statistical backtesting has proven the indicators are frequently a better predictor than street-mean estimates, while also having the advantage of substantial lead time.

Dataset

Trade nowcasting indicators provide valuable insights into international trade and industrial production. Our data partner employs big data and predictive analytics techniques, and more than 25,000 times series, to forecast trade balance and industrial production statistics. The data is also available at the level of an individual shipping port, and bespoke feeds can be provided.

Geography	Coverage	Mapped to Tickers	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
Global	12 Countries e.g. USA, CAN, JPA, GB, GER, TAI	No	Since 2010	Daily	Daily	1 day	API, CSV

Backtesting/Significance

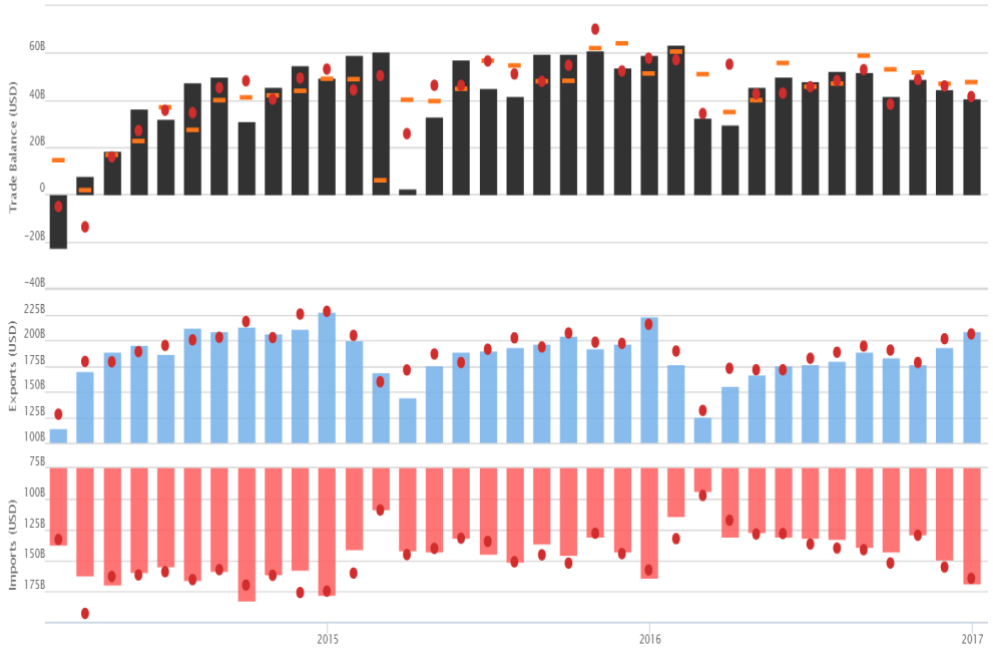
Statistical backtesting has proven the indicators are frequently a better predictor than street-mean estimates, while also having the advantage of substantial lead time. Over the last 3 years, the indicators were better than street estimates 67% of the time.

Case Study

Considering China's trade balance, Figure 64 below shows monthly values from February 2014 onwards. Red dots show the trade nowcasting forecasts, while orange dashes indicate street estimates.

In January 2017, it was reported that China's trade remained sluggish in December 2016 with exports decreasing 6.1% YoY. Imports, on the other hand, increased by 3.1% YoY vs. 6.7% YoY in November 2016.

Figure 64: China's Trade Balance



Source: Trade Nowcasting Indicators

35. Macro > Discretionary (long-term) > Trade Data

Key Takeaway

The South Korea real-time export data accurately tracked overall China exports.

Dataset

This vendor delivers high frequency and comprehensive South Korea export data. Preliminary export data includes total volume and value of all products, at all country destinations that are exported from South Korea. The dataset is created by aggregating and analyzing customs declaration forms.

Dataset Overview:

• History: Since 2003.

• Geography: South Korea export destinations.

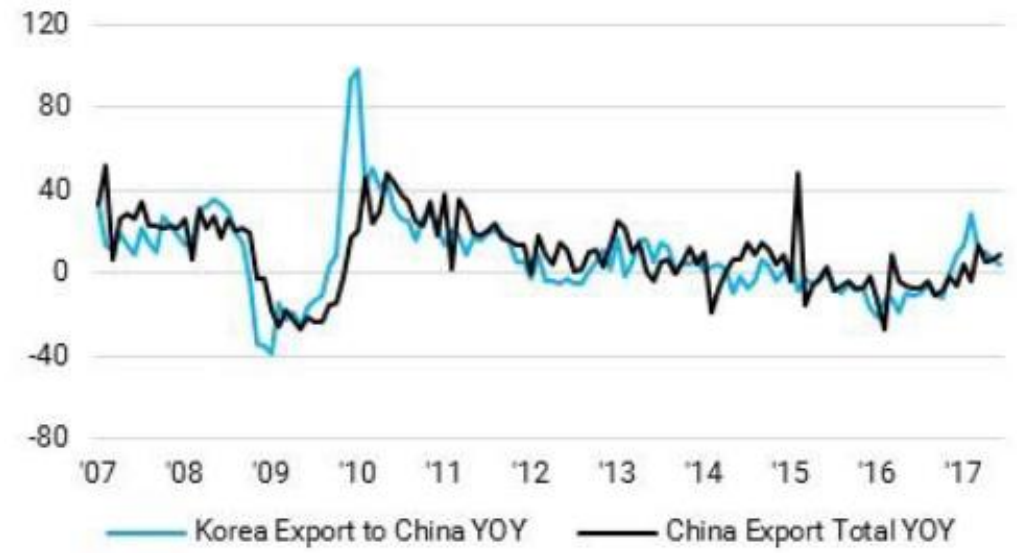
• Delivery: CSV.

• Frequency: Weekly.

Case Study

The South Korea real-time export data accurately tracked overall China exports in the 2007-2017 period.

Figure 65: South Korea Export vs China Total Exports



Source: Trade Data Provider

Disclaimer

This case study is directly from the vendor and Eagle Alpha has not backtested the data.

36. Macro > Discretionary (Long-Term) > Employment

Key Takeaway

Using payroll processor microdata improves forecast accuracy of an aggregate labor market activity measure.

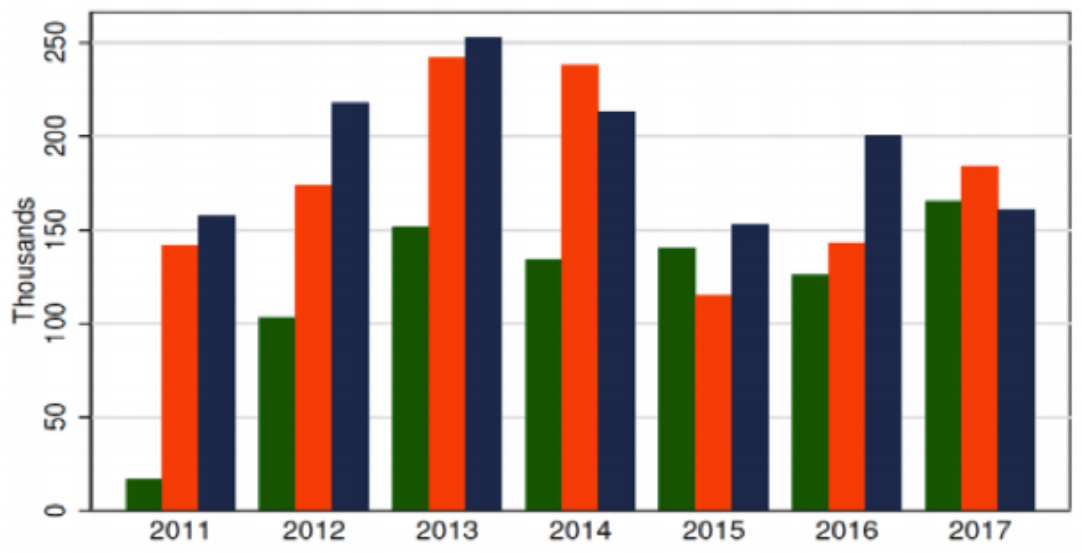
Dataset

The vendor is one of the largest providers of business processing and cloud-based solutions – including payroll, talent management, human resource management, benefits administration and time and attendance – to employers and automotive dealerships around the world. History: since January 2014.

Case Study

Members of the Federal Reserve Board evaluated value of the payroll processor microdata and how it could be used to measure aggregate labor activity. They highlighted that the timeliness and frequency of the payroll data improved forecast accuracy for the official figures (Current Employment Statistics published by BLS, the Bureau of Labor Statistics). Some of the results are presented below.

Figure 66: Real-Time Data (Blue) vs First BLS Print (Green) vs Latest BLS Figures (Orange)



Source: BLS, Employment Data Provider.

37. Macro > Discretionary (Long-Term) > Sentiment

Key Takeaway

Quantitative metrics of emotional content in market narratives may complement other indicators and analysis in helping to gauge systemic risk.

Datasets

Three data sources were analysed: 1) Internal Bank of England daily commentary on market news and events, i.e. 26 documents per month on average from January 2000 until July 2010; 2) Reuters' news wire articles in the United Kingdom, consisting of over 17 million English news wire articles; and 3) Broker research reports, which are from 14 brokers from June 2010 until June 2013, approximately 100 documents per month.

Case Study

The Bank of England researchers analysed unstructured text-based market data to explore how narratives and sentiments drive developments in the financial system. They concluded: "We find that changes in the emotional content in market narratives are highly correlated across data sources. They show clearly the formation (and subsequent collapse) of very high levels of sentiment – high excitement relative to anxiety – prior to the global financial crisis."

The researchers used a word count methodology (see Figure 67 below), which classifies words as representing anxiety or excitement, to calculate an emotional statistic of sentiment (Relative Sentiment Shift or RSS). RSS equals the difference between excitement words and anxiety divided by the total number of characters in a text.

Figure 67: Word Count Methodology Used

Table 1: Emotion dictionary samples

Anxiety		Excitement	
Jitter	Terrors	Excited	Excels
Threatening	Worries	Incredible	Impressively
Distrusted	Panics	Ideal	Encouraging
Jeopardized	Eroding	Attract	Impress

Source: The Bank Of England

Further robustness tests showed that the sentiment metrics have some predictive power for standard aggregated measures of consumer confidence, market volatility, and some relevant but more theoretic measures of uncertainty from the literature exploiting text-based information. Furthermore, their measures also contain useful information in explaining economic data in a simple VAR model. Therefore, the authors of this paper believe that these quantitative metrics of emotional content in market narratives may complement other indicators and analysis in helping to gauge systemic risk.

38. Macro > Discretionary (Long-Term) > Trade Nowcasting

Key Takeaway

Trade data was used to track the surprising macro-economic turnaround for Brazil in 2016.

Dataset

Trade nowcasting indicators provide valuable insights into international trade and industrial production. Our data partner employs big data and predictive analytics techniques, and more than 25,000 times series, to forecast trade balance and industrial production statistics. The data is also available at the level of an individual shipping port, and bespoke feeds can be provided.

Geography	Coverage	Mapped to Tickers	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
Global	12 Countries e.g. USA, CAN, JPA, GB, GER, TAI	No	Since 2010	Daily	Daily	1 day	API, CSV

Backtesting/Significance

Statistical backtesting has proven the indicators are frequently a better predictor than street-mean estimates, while also having the advantage of substantial lead time. Over the last 3 years, the indicators were better than street estimates 67% of the time.

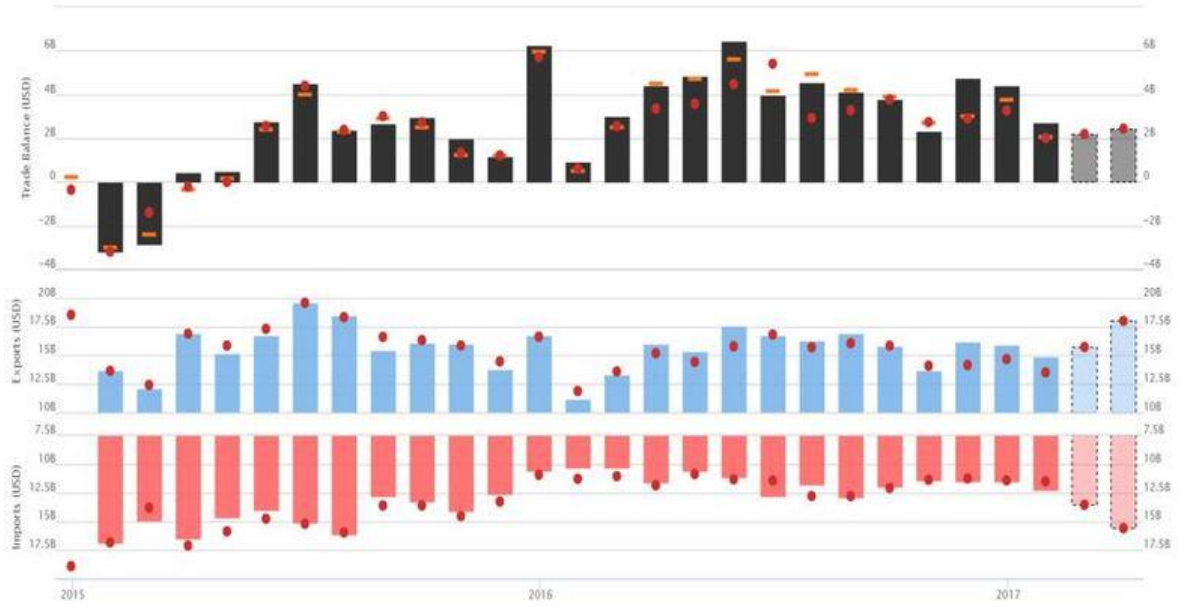
Case Study

One of the most surprising macro-economic turnarounds in 2016 occurred in Brazil. In April 2016, Brazil's economy was in a free-fall, exacerbated by political turmoil. However, the trade nowcasting indicators showed that there were hints of moderation in the steep drop in shipments into the country.

The nowcasting indicators appeared to be forming a bottom and pointed to signs of economic relief in May 2016. In July, the data was used to increase Brazil's industrial production forecast and in August the conclusion was made that the country's economy was exhibiting a turnaround.

Finally, in October 2016, data on Brazilian consumer goods imports signalled the return of the Brazilian consumer. In Figure 68 below, red dots show the trade nowcasting forecasts, while orange dashes indicate street estimates.

Figure 68: Brazil's Trade Balance



Source: Trade Nowcasting Data Provider

The Brazilian stock market surged as a result, increasing 66% in 2016 as measured by MSCI Brazil index.

39. Credit > Discretionary (Long-Term) > Pricing Data

Key Takeaway

Online property listings data was used to research buy-to-rent investors which led to the subsequent clamp down on mortgage lending.

Dataset

The case study below is based on an online property listings dataset that covers 80% of UK sales. This dataset is more timely than any publicly available dataset. Our partner's data can be actionable in various ways, such as analysing property deal flow and volume, tracking liquidity of the property market, tracking the liquidity and availability of secure consumer credit, analyzing consumer confidence around envisaged indebtedness, and advanced indicator of strength or weakness in housing related consumer discretionary stocks as well as telecom and utilities.

Geography	Coverage	Mapped to Tickers	History	Collection Frequency	Delivery Frequency	Lag Time	Delivery Method
U.K.	80% of U.K home sales	No	Since 2005	Daily	Daily	1 Day	FTTP, CSV

Case Study

Bank of England analysts demonstrated their research in several papers: 'How much do investors pay for houses?' (September 2015)[26], 'Five facts about buy-to-let' (July 2015)[27], 'Chance favours the prepared mind' (July 2015)[28]. These articles shed a light on the Bank's methodology and how various micro datasets were linked in order to gain valuable insights into UK housing market trends.

Buy-to-rent investors – known as buy-to-let (BTL) in the UK – were becoming large players in the UK housing market. Housing stock held for private renting went from 9% of the total stock in 2000 to 19% in 2013 (Bracke, 2015). The Bank of England decided to investigate whether BTL investors drove the housing prices up and could lead to macro instability.

The online property listings dataset was combined with Product Sales Data (PSD) and Land Registry's Price Paid data. The PSD data provided by the Financial Conduct Authority was used to gather details of loans for house purchases.

The combination of these datasets allowed the Bank of England to conduct a more precise analysis and get closer to the understanding of the individual agents' decision-making. Some valuable insights are presented below.

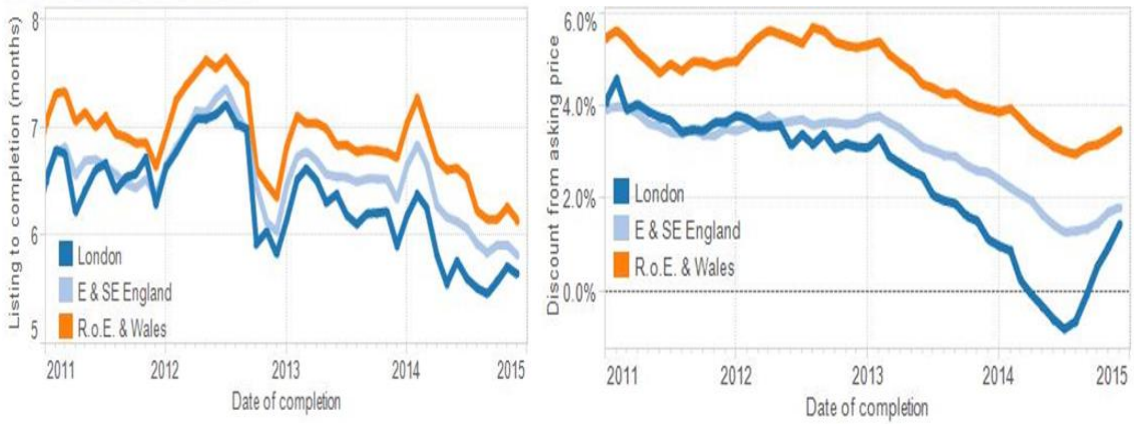
A decrease in average time from listing to completion of a sale highlighted increased purchasing activity.

[26] Source: 'How much do investors pay for houses?', September 2015

[27] Source: 'Five facts about buy-to-let', July 2015

[28] Source: 'Chance favours the prepared mind: What linked micro data can tell us about the housing market', July 2015

Figure 69: Months from Listing to Completion *Discount from Last Asking Price*



Sources: Online Property Listings Dataset, Land Registry Price Paid, ONS Postcode Directory

Buyers were able to negotiate and obtain 4-5% discounts in 2011-2012. However, the analysis showed that buyers lost some of the bargaining power as the speed of transactions increased. In London, the average discount even turned negative at one point.

Bracke (2015)[29] concluded: "The data show that BTL investors can accelerate the time it takes to sell a property, and BTL discounts are the implicit compensation for this contribution. However, investors' ability to `grease the wheels' of the housing market becomes limited when the market is already performing well. This is precisely when financial stability concerns become most important."

The Bank of England then used the above analysis to further investigate the mortgage-financed part of the housing market as half of all BTL transactions were supported. It then decided to take action and clamp down on BTL lending: "The PRA, the Bank's regulation arm, is concerned that changes to mortgage interest tax relief for landlords will strain buy-to-let borrowers and that only a few lenders include this risk when assessing mortgage applications."

[29] Source: 'How much do investors pay for houses?', September 2015

40. Credit > Discretionary (Long-Term) > Credit Risk Data

Key Takeaway

Testing showed that the probability of bankruptcy within 12 months ranges from 10% to 50% when a company gets a stress score of "1".

Dataset

Credit score data provider delivers predictive stress scores for credit, supply chain and financial professionals. Credit ratings are prepared for over 57,000 public companies worldwide and can be used to analyze equities with high bankruptcy risks.

Aggregate crowd-sourced usage data from the vendor's subscribers, credit managers and supply chain professionals from Fortune 1000 companies, is incorporated during the construction of stress scores. The provider found distinct behavioural patterns when its subscribers are concerned with certain companies and investigate them more closely.

Backtesting/Significance

Monthly credit scores have been proven to be highly accurate when predicting corporate failures. Testing showed that the probability of bankruptcy within 12 months ranges from 9.99% to 50% when a company gets a score of "1".

Case Study

On August 31st 2016, Hanjin Shipping, the world's seventh-largest container carrier at the time, filed for bankruptcy protection as it could not renegotiate its debts. On September 12th 2016, it was estimated[30] that $14 billion worth of cargo was stuck aboard Hanjin ships and the company lost a third of its market value in two weeks.

On 21st September 2016, the vendor published a case study with the post-filing analysis of Hanjin Shipping's bankruptcy. Figure 70 below shows that the credit score for Hanjing dropped to "2" in December 2015 and "1" in February 2016 well in advance of the bankruptcy filing.

[30] Source: 'A By-the-Numbers Look at Hanjin Shipping's Collapse', September 2016

Figure 70: Monthly Average Credit Scores

Business Name	2015 SEP	2015 OCT	2015 NOV	2015 DEC	2016 JAN	2016 FEB	2016 MAR	2016 APR	2016 MAY	2016 JUN	2016 JUL	2016 AUG	2016 SEP
2GO Group Inc	6	6	7	7	7	6	6	7	7	7	7	7	7
Abetrans Ltd.	3	3	3	3	3	3	3	3	3	3	3	3	3
AP Moeller Maersk A/S	9	9	9	10	9	9	9	9	9	9	9	9	10
Arpeni Pratama Ocean Line Tbk PT	2	2	2	2	2	2	2	2	2	2	2	2	2
Associated British Ports Holdings Plc				9	9	9	9	9	9	9	8	8	8
Atlantska Plovidba dd	4	4	4	4	4	4	4	4	4	4	4	5	5
Avance Gas Holding Ltd	7	8	8	8	8	7	7	7	5	4	3	3	3
Barska plovidba A.D. Bar	4	4	4	4	4	4	4	4	4	4	4	4	4
Bollore SA	6	6	6	6	6	6	5	5	5	5	4	5	5
Buana Listya Tama Tbk PT	6	6	6	6	6	6	4	4			5	5	6
China COSCO Holdings Company Limited	4	4	4	4	4	4	4	3	2	2	2	3	4
China Merchants Energy Shipping Co., Ltd	5	5	6	6	6	6	6	6	6	5	6	6	6
Chu Kong Shipping Enterprises(Grp)Co Ltd	7	7	7	7	7	7	7	7	7	6	7	8	8
Companhia Doca de Imbituba	3	3	3	3	3	3	3	3	3	3	3	3	3
Express Kenya Limited	3	3	3	3	3	3	3	3	3	3	3	3	3
FUSHIKI KAIRIKU UNSO CO., LTD.	5	5	5	5	5	5	5	5	5	5	5	5	5
Hanjin Shipping Co Ltd	3	3	3	2	2	1	2	1	1	2	1	1	1
Kirby Corporation	9	9	9	9	9	9	10	10	10	10	10	9	9
OT Logistics SA	6	6	6	6	6	6	6	6	6	6	6	6	6
South Logistics JSC	7	7	7	7	7	7	7	7	5	5	5	5	5

Source: Credit Risk Data Provider

Disclaimer

This case study is directly from the vendor and Eagle Alpha has not backtested the data.

Section 6: Overview of Eagle Alpha

Since 2012 Eagle Alpha has provided a full service solution that enables asset managers to obtain alpha from alternative data.

Eagle Alpha's offering has continually evolved since 2012 based on client feedback. There are four parts to the offering – see Figure 71.

1. Data Sourcing keeps clients on top of all relevant raw and semi-processed alternative datasets worldwide.
   - The Data Sourcing offering includes a database (online and API), exclusive datasets, an advisory service, data showcase events and roadshows.
2. Data Analytics enables clients to interpret alternative data using tools, indicators and analyst insights.
   - The Data Analytics offering includes tools, indicators, reports and analyst access.
3. Bespoke Projects enables clients to commission curated work for specific research questions.
   - Bespoke projects can include data acquisition, data engineering, data analysis and predictive modelling.
4. Data Forum enables clients to keep on top of the evolving alternative data space.
   - The Data Forum includes meetings (online and offline) and conferences, working groups, an advisory service and a knowledge center.

Since 2012 Eagle Alpha has built a 25 person team that is modelled on comparable teams at larger asset managers that have embraced alternative data. Our product team has four groups:

1. Data Sourcing: our data hunting team hunts (proactively and reactively) for interesting alternative datasets.
2. Data Analysts: our data analysts, who have buyside and sellside experience, understand asset classes and investigate strategies.
3. Data Scientists: our data scientists have strong quantitative skills for data exploration and backtesting.
4. Engineers: our engineers do data engineering and deliver our solution to clients.

Asset management clients include quantitative funds, discretionary hedge funds, long only funds, soverign wealth funds and private equity funds. We are uniquely positioned to work with clients that are starting to integrate alternative data as well as sophisticated users that use Eagle Alpha to complement their in-house capabilities.

Eagle Alpha is a recognized leader in the alternative data space. We have been heavily referenced in alternative data primers by several firms e.g. Citi, JPMorgan, BoA and Deloitte.

Our vision is to be the 'go-to' firm for the alternative data needs of asset managers.

Figure 71: Overview of Eagle Alpha

How do you want to consume alternative data:

1. Raw and/or semi-processed data?
2. Use tools, indicators and analyst insights?
3. In a curated form?

Data Sourcing

Database (Online & API), Exclusive Datasets, Advisory Service, Events & Roadshows

Data Analytics

Tools, Indicators, Reports, Analyst Access

Bespoke Projects

Data Acquisition, Data Engineering
Data Analysis, Predictive Modeling

How can you keep on top of the evolving alternative data space?

Data Forum

Meetings & Conferences, Working Groups, Advisory Service, Knowledge Center

Since 2012 Eagle Alpha has provided a full service solution that enables asset managers to obtain alpha from alternative data

Eagle Alpha