Big Brother is Watching You! How Can Alternative Data Help Traders to Beat the Market

Introduction

Information has always been one of the main sources of alpha on the market. Centuries ago just having first-hand information about the outcome of a war could very well constitute an important edge over others, the famous speculation of Nathan Rothschild on the Battle of Waterloo comes to mind. As time passed and technology improved, innovative sources of the past became obsolete. Stock prices and data on corporate financials, which once constituted an important edge on the market, are now easily and freely accessible.

Similarly, in 2010 the ultra-low latency market access provided specialised trading companies with an incredible edge while nowadays it is seen almost as a prerequisite to enter the market rather than a competitive advantage, and the level-playing field has been largely levelled up.

In recent years the growth of computing power and interconnectedness led to the explosion of unstructured data, originating from the most disparate sources such as smartphones and cars.
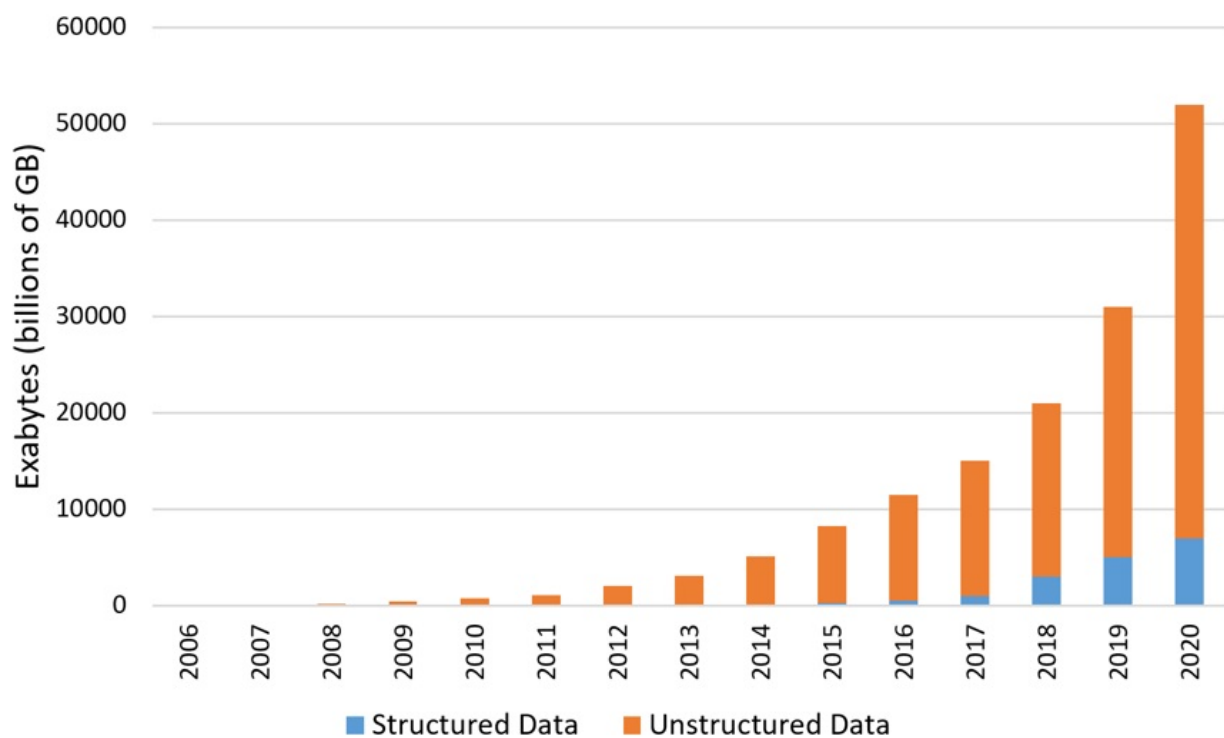
*Source: Electronic Engineering Times*

Big data has been called the new oil, it comes in all forms and shapes and takes a lot of effort to process it in something useful. For this reason, data science is now one hottest field around and companies are rushing trying to get an edge over the competition.

It's in this environment that the so-called alternative data is thriving.

Alternative data refers to data coming from unconventional sources which hedge funds and other institutional investors can use to obtain innovative sources of alpha. As an example, these datasets can include credit card transactions, social media feeds and geospatial imagery. Given the complexity and the volume of this data, more and more funds are hiring teams of data scientists to keep track, analyse and process it.
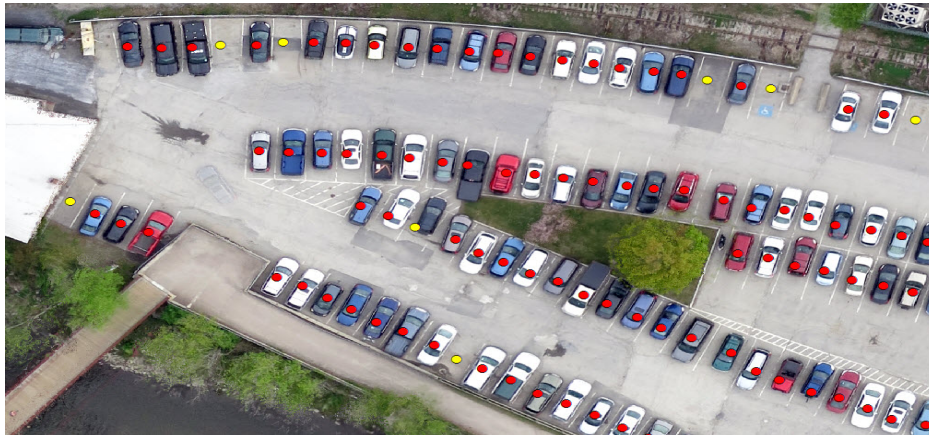
Alternative data goes mainstream

■ Number of alternative data analysts   ■ Number of investment funds employing alternative data analysts
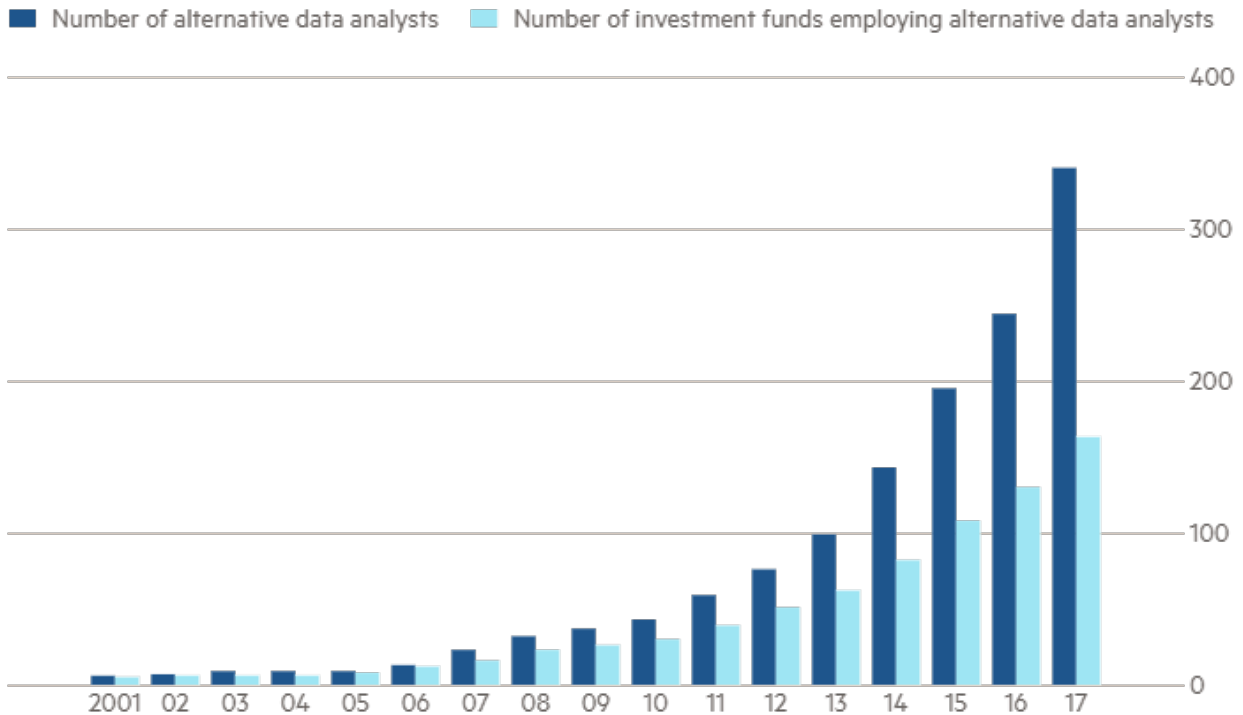
Source: AlternativeData.org

© FT

Types of Alternative Data

The alternative data industry has now reached $200 million in size, but some general categories of data and data providers can be discriminated among this huge pool. Indeed, alternative data can be more or less expensive and its availability might be exclusive (hedge funds can make agreements with providers to be the exclusive beneficiaries) or can reach a wide range of users. Nevertheless, it is estimated that 24% of hedge funds implement alternative data technologies, with the most common categories described below.

The most well-known branch of alternative data is perhaps the satellite imagery: companies get data from satellite data vendors and such sources are analysed by advanced analytics tools and machine-learning software to get to final outputs. These may provide insights on traffic outside malls, the status of crops in the countryside or even keep track of the number of factories under construction in rural China. It is subsequently possible to forecast the mall's revenues, the price of agricultural commodities or the Chinese economic activity. FeatureX and Descartes Labes are examples of companies quickly developing these technologies.

Another quite powerful indicator originating from alternative data sources is the oil depot floating tank shadow height: by measuring the shadow on one side of the tank you can find a reference point for the total height of the tank, and subsequently infer on the amount of oil contained in the tank. It helps traders to anticipate the oil reserves amount and profit from taking a position based on data not publicly available on the market. Other uses involve the study of agriculture and crops to predict agriculture commodities' prices, and estimation of mining yields.

This can be used along with the geo-location services, i.e. location data derived from mobile devices, which helps to understand the impact of promotions by identifying the number of people flowing to a specific mall. Similarly, cross brand loyalty and regional idiosyncrasies may be identifiable.

Another valuable source of information for hedge funds are consumer transactions (especially credit card payment data) along with customer relationship management one. Point72 has acknowledged of using satellites and credit cards data to forecast the number of clients and their expenditures in stores and malls. This category ranges from merchant (airlines, services provider) to product level (beverages and cars) transaction data. Credit card data are not only widely used only by asset managers to understand the quarterly level of economic activity sector by sector, but also lenders often use such data to come up with better indicators than the traditional credit score to evaluate clients' financial stability. Data providers usually anonymize the data they gather from credit- and debit-card transactions and sell some of it directly to hedge funds and to research firms. These usually represent only subsets of single merchant data. Interestingly enough, this alternative data method caused increased volatility in retailers' stocks, the most affected sector. As a matter of fact, credit card data has often helped insiders to correctly forecast

earnings. However, in the cases of credit card data pointing at the wrong figures, the market has often witnessed wider-than-normal swings during the day of earnings, as institutional investors re-position themselves.

The third most sought-after category is internet/social media data both to understand the overall sentiment of the public towards events (elections, scandals…) and to be the first to know of breaking news. Sentiment scoring may be applied to investor commentary, consumer attitudes toward products and brands, or mainstream news feeds. Bloomberg was an early-developer of such technologies, as it launched in 2015 the Twitter sentiment count, which analyses the number of positive and negative tweets on a specific company, so that it is possible to notice how these have an impact on the underlying stock. It simply searches for tweets which directly involve the company and discern positive from negative ones. Finally, it returns the count on the terminal. These tools prove particularly useful to evaluate product launches, trends and how the ESG (environmental, social and governance) aspects of a firm are viewed by the public.

One further macro-category of alternative data refers to weather: alerts to breaking news from major news wires or social media sources allow traders to react before news is fully discounted in asset prices. In such a highly volatile market like the power market getting to know as soon as possible the current weather situation is crucial for the trader to make profits: to do so a combination of numerical weather prediction (NWP) models and statistical ensemble of forecasts. It is then possible to take a position on options and futures concerning specific cities (CME has instruments focusing on major US cities, London, Amsterdam and a few others).

Other alternative data types involve online search: 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. Nevertheless, complex data science techniques are needed and still developing to separate the number of searches that are most likely to end up in a purchase of a product from noise. Also pricing data is becoming more readily available and it might be a useful hint to understand inventory levels. An elevated number of discounted products might mean that the firm has necessity to free up inventories, hence increasing markdown sales. Similarly, it might also indicate a deflationary economy, thus enabling traders to speculate on next central bank's interest rate movements. Finally, one last major source of alternative data are reviews and ratings: favourable reviews will increase sales and vice versa. Also, app ratings can help understand the overall level of customer satisfaction with the online services of a shop. It is worth mentioning also how the reviews of business customers on B2B platforms like Alibaba are

increasingly becoming an important influence for the market value of firms, in particular as B2B platforms grow in importance.

Here is a list of the current major players of the industry, but the list in continuous evolution as the alternative data industry evolves and changes:

Satellite Data	Credit Card Data	Social Media Data	Weather Data	Location/ Foot Traffic	Data Aggregators
Descartes Labs	Earnest	7park Data	Aclima	Airsage	Discern
Feature X	Envestnet Yodlee	App Annie	Climate Corporation	Foursquare	Eagle Alpha
Orbital Insights	Second Measure	Dataminr	Understory	Placed	Quanton Data
Planet Labs	Ynext Data	Datasift		Placemeter	Sentieo
RS Metrics		Selerity		StreetLightData	
Spire		Yipit Data			

Strategy Cases Examples

Here follows a short list of examples on how alternative data strategies were implemented and exploited in the market.

Thasos Group is an alternative data provider who was co-founded by one of the most powerful data scientists in the world, Alex Pentland. Thasos provides real-time locations from mobile phones which can then be used to gather insights on the performance of business and markets. On August 2017, Amazon acquired Whole Foods and immediately applied price reductions, and in order to quantify this new measure, Thasos platform was used. It was concluded that the foot-traffic to Whole Foods after a week of the price reduction increased by 17%. Thasos could also observe from which supermarkets the customers were coming from, 24% from Walmart, 16% Kroger and 15% Cosco. The price reduction only attracted the wealthiest customers of the competing stores, the lower demographic ones and the ones that had to drive the furthest to reach a Whole foods store were not as incentivized.

Eagle Alpha provides its solutions in three different parts on which the client can choose from, these are: Data Sourcing which consists of databases with raw or semi-processed data, Data Analysis which delivers tools and indicators and Bespoke project which gives curated data. Eagle Alpha has consistently proven that customers have achieved alpha with its databases. A cross-sectional analysis of a partner dataset consisting in anonymized purchase data of 2 million shoppers was scanned via email receipts. They were able to turn unstructured email data into a normalized dataset, which was also granular and contained 53 product categories. One of their client, JP Morgan, was able to achieve 16.2% annualized return and a Sharpe ratio of 1.13. Eagle Alpha could also predict that Just Eat would have stronger than expected revenues in H2 2015 and H1 2016 with its partner platform which provides consumer transaction data from UK's largest personal financial assistant (an app/web platform).

Dataminr presents breaking information and events which have a high impact in instants before these reach the news. It has solutions for Corporate Security, Finance, the Public Sector, News, and PR / Communications. On the 2nd of March 2018, groups of extremists attacked Burkina Faso's capital. The attack injured 80 and killed 8. It was eyewitnesses that first reported the event on social media and as Dataminr uses proprietary algorithms to screen public tweets, it delivered the earliest alert.

Looking Ahead

Given the constant gathering of data which is then used by companies, some concerns arise regarding privacy and the not-so-clear distinction between informational edge and insider trading. Personal information should be scrubbed from datasets which are sold to outsiders but there is not a clear standard in the industry. Although buying and selling data is legal in many countries the recent public outcry at the Facebook-Cambridge Analytica scandal might very well spark more privacy-oriented legislation, especially in countries with traditionally weak data protection policies such as the United States.

Other concerns regard the legal aspects of alternative data and particularly its relationship with insider trading. Although many data sources are effectively public, others aren't and instead are often sold on exclusive basis, potentially raising regulatory concerns. A recent report by Integrity Research Associates cites the case of a successful prosecution by the SEC where two analysts obtained credit card transactions without the consent of the owner of the data and were thus prosecuted for insider trading. The report suggests that the introduction of industry-wide compliance standards might reduce legal risk and compliance costs as well as shielding the industry from adverse media and regulatory scrutiny.

Finally, it's important to stress that although alternative data is currently an innovative and unconventional source of alpha, eventually, like other sources, will become a standard in the industry and will cease to be "alternative". As this is still quite a new sector, there is a lot fragmentation in the market with many companies specialising in a niche of the market. We forecast a future consolidation of the market, in which companies will provide a wider range of services, and increasing economies of scale, by M&A. A consolidation of the industry is expected to take place as the industry grows more and more.