

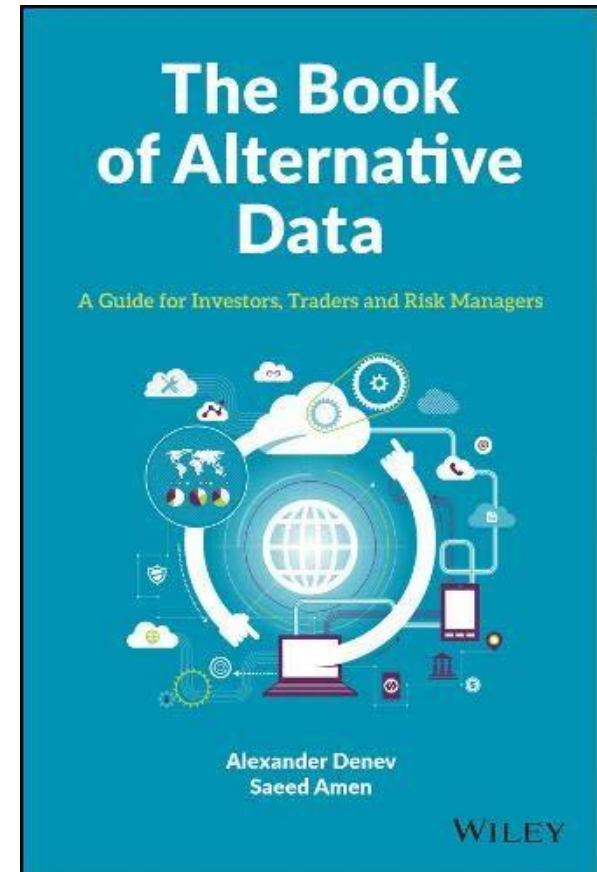
ALTERNATIVE DATA: A GUIDE FOR INVESTORS, TRADERS AND RISK MANAGERS

WEBINAR: ALTERNATIVE DATA USE
IN ASSET MANAGEMENT

FEBRUARY 2022

THE BOOK OF ALTERNATIVE DATA

- Co-authored by Alexander Denev and Saeed Amen
- **The Book of Alternative Data** (published Wiley in mid 2020)
- You can find it on Amazon



WHAT IS ALTERNATIVE DATA?

- **Common properties**

- Less commonly used by market participants
- Tends to be more expensive
- Often outside financial markets (is tick data “alternative”?)
- Shorter history
- More challenging to use

- **“Exhaust data” a byproduct of other processes**

- Digital footprint from individual and corporate activity
- Resulted in a rapid rise in the number of alternative datasets
- Can provide an additional revenue stream for those who collect “exhaust data”

THE V'S OF BIG DATA

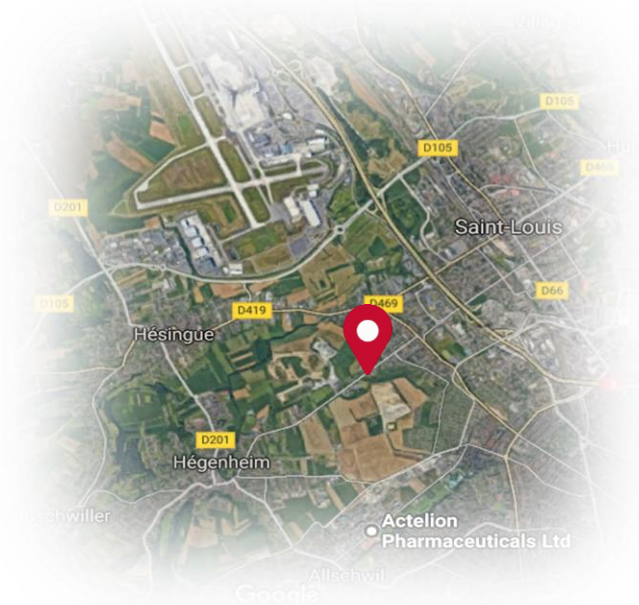
- **Volume (increasing)** — lots of data
- **Variety (increasing)** — not just numerical data, can be text, image, video etc.
- **Velocity (increasing)** — speed that data is being generated
- **Variability (increasing)** — inconsistencies in the data
- **Veracity (decreasing)** — difficult to tell if accurate (e.g. social media)
- **Value (increasing?)** — business value of the data

TYPES OF ALTERNATIVE DATA

- **Satellite/aerial photography**
- **Location data**
 - mobile phones
 - apps
- **Text**
 - Web
 - Social media
 - News
 - Internal data
- **Consumer transactions**
 - Credit card transactions
 - E-mail receipts
- **Corporate**
 - Supply chain
 - Internal metrics
- **Market**
 - High frequency tick
 - Flow data
- **Crowdsourced data**
 - Scouts on ground
 - Analyst estimates
- **And much more!**
- **Have some case studies later and in our book!**

DATA IN THE BLIND SPOT

Hedge funds Och-Ziff, Eton Park and Elliott Management are said to have tracked J&J's private jet landing near pharmaceutical company Actelion's headquarters in Switzerland. Days later J&J announced a \$30 billion acquisition of Actelion, resulting in \$300M in upside for these three funds.



ALTERNATIVE DATA & INVESTMENTS CASE STUDIES

Several clear case studies have emerged demonstrating the value of analytics in combination with alternative data applied to the investment process



ON-LINE PRICE = INFLATION

Global FSI Firm employs technology to track prices of 5 million products on-line to understand price shocks and monitor shifts in inflation across 70 countries¹



APP + CREDIT CARD = PERFORMANCE

Hedge Fund looks at combination of alternative data including credit card transactions, geo-location, and app downloads to analyze burger chain performance²



SOCIAL + SEARCH = EARNINGS

\$90B AUM Global Asset Manager mines search engine data combined with social-media data to predict results of corporate events like quarterly earnings³



MOBILE FOOT TRAFFIC= ECONOMY

Hedge Funds using location data pulled from mobile devices to predict outlook on economy and REIT values⁴



SATELLITE + SHIPS = MISPRICED SECURITY

Hedge fund using satellite intelligence on ships and tank levels to identify upcoming impact to oil producers and commodity prices⁵



WEB + TWITTER = MARKET MOVING EVENTS

Data provider using 300M Websites, 150M Twitter feeds in combination with analyst presentations and FactSet reports to measure rise up media food chain (e.g. blogs to newswire) to highlight potentially market moving events⁶

Doug Dannemiller, Alternative Data and Collective Intelligence Investing: Risks of Adoption, Deloitte

1. Innovative Asset Managers, Eagle Alpha

2. "Foursquare Wants To Be The Nielsen Of Measuring The Real World," Research Briefs, CBInsights, June 8, 2016.

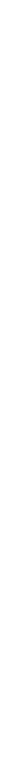
3. Simone Foxman and Taylor Hall, "Acadian to Use Microsoft's Big Data Technology to Help Make Bets," Bloomberg, March 7, 2017.

4. Rob Matheson, "Measuring the Economy With Location Data," MIT News, March 27, 2018.

5. Fred R. Bleakley, "CargoMetrics Cracks the Code on Shipping Data," Institutional Investor, February 04, 2016.

6. Accern website

THE MARKET PARTICIPANTS



ADDRESSING MARKET CHALLENGES

Quantitative investment strategies and vendor solutions with alpha generation capabilities are becoming critical component to the return of the buy and sell side's ROE to pre crisis levels

Sophisticated Quants	Traditional Quants	Traditional Investors	Fintechs
<i>Systematic/quant Investors, typically building their own analytics</i>	<i>Interested in derived analytics and more intuitive solutions</i>	<i>Most intuitive solutions needed. Limited technology and programming capability</i>	<i>Sophisticated but ultra small scale with a focus on highly scalable business models</i>
Who:	Who:	Who:	Who:
<ul style="list-style-type: none">• Hedge Funds• Sophisticated Buy Side Firms	<ul style="list-style-type: none">• Large Sell Side (GSIBs)• Traditional Buy Side Firms	<ul style="list-style-type: none">• Smaller Sell Side (DSIBs)• Discretionary investors	<ul style="list-style-type: none">• Alternative Data Providers• Signal Factories
Key Challenges:	Key Challenges:	Key Challenges:	Key Challenges:
<ul style="list-style-type: none">• Access to good quality raw data or to curated alternative data• Maintaining access to cutting edge technology and algorithms• Completeness of data	<ul style="list-style-type: none">• Reducing technology costs associated with efficient research tools• Retention and expansion of innovation talent	<ul style="list-style-type: none">• Reducing technology costs associated with efficient research tools• Building/maintaining an edge against passive benchmark returns	<ul style="list-style-type: none">• Simplified access to data• Ability and agility to scale• Support of cutting edge algorithms and alternative data sets
Customer Needs:	Customer Needs:	Customer Needs:	Customer Needs:
<ul style="list-style-type: none">• Co-location of analytics and data• Simplified access to data and computation• Simplified, but bespoke, data access	<ul style="list-style-type: none">• Simplified access to data and computation• Curated Signals• Simplified, but bespoke, data access	<ul style="list-style-type: none">• Simplified access to data and computation• Curated Signals• Sophisticated, but low maintenance/build cost analytics platforms• Elastic access to analytics and associated data science talent	<ul style="list-style-type: none">• Simplified access to data and computation• Simplified, but bespoke, data access• Sophisticated, but low maintenance/build cost analytics platforms• Marketplace creation

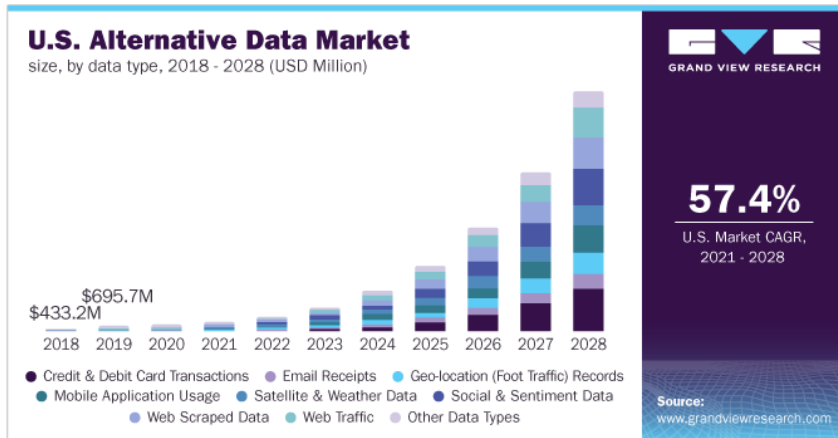
GETTING TO GRIPS WITH NEW DATA SOURCES AND TECHNIQUES

Investors are increasingly spending on alternative data but building data science and engineering teams, and the associated analytics platforms to fully harness such diverse data, remains a significant barrier for all but the largest firms.

Market Trends

Buy side spend on alternative data has increased over the previous 3 years and is expected to continue to grow

- Poor active investment performance is driving shift to passive products and fee compression
- Active investing strategies are starting to require more diverse data to generate strong alpha and beta predictive signals
- Savings from bundling of data streams are not currently possible due to the segmentation of the data providers market but are becoming highly desirable



Barriers to Entry

Setting a data science/engineering team can be both expensive and time consuming:

- Cost of an alternative data team could start at \$1.5 - \$2.5m
- The technology and methodology needed for integrating different datasets further increases costs
- Processes that are not optimally engineering can also lead to failures and costs
- Data quality is an issue in which significant resources are invested

	Entry Level Salary	Bonus
Data Analyst	80k – 100k	~25%
Data Scientist	80k – 120k	~40%
Data Scout	70k – 90k	~15%
Data Engineer	80k – 110k	~30%
Head of Data	250k - 1m	~100%

Full Data Team

- 1 Data Engineer
- 3 Data Analysts
- 1 Data Scientist
- 1 Data Scout
- 1 Head of Data

Alt. Data FTE Comp: **\$1.5m - 2.5m**

Note: Top funds spend over \$10m on alternative data FTEs

Source: Alternativedata.org

CHALLENGES



LEGAL QUESTIONS

- **Before buying data, we need to be aware of the legal aspects:**
 - Can the data be sold? (e.g. GDPR issues and consent)
 - Have the personal details been properly scrubbed?
 - Does the data need to be aggregated before being sold to “blur” it?
 - Are there issues for “exclusive” datasets?
 - Very important for sellers to be aware of the legal aspects (as well as buyers), must investigate beforehand
 - Issues will vary between datasets

DATA CHALLENGES

- **Entity matching**

- Matching to traded assets (e.g. iPhone to Apple)

- **Missing data**

- Data can be sparse, how can we fill (averages?)

- **Structuring the data**

- Converting unstructured data, often images and text into a more structured form, often ultimately into a time series of numerical data

- **Anomalies**

- Data which deviates substantially from what is expected, e.g. outliers in tick data

RISK EXPOSURE

Alternative data carry greater risk than traditional data and these datasets may also introduce newer risk types



Data risk: Firms may face these types of data risks due to immature risk control processes at data providers

- **Data provenance risk:** Violation of the terms and conditions from the data originator while scraping websites
- **Accuracy/validity risk:** Data may prove unreliable or produce an inaccurate trading signal
- **Privacy risk:** Personally identifiable information could be included in a dataset received from a source
- **Material non-public information (MNPI) risk:** Receipt of a dataset containing MNPI could result in risk events



Model risk: The potential of new data sources to impact the investment models and perhaps decision making, if:

- The data is incorporated in the model incorrectly
- The trading signal generated is irregular or inconsistent under certain conditions
- The output of the model is improperly linked to the trading process



Talent risk: IM firms may face the following risks due to the rise in demand for data science and advanced analytical skills to process alternative data:

- Loss of intellectual capital through talent turnover
- Impact on alternative data utilization ability due to delayed training for existing employees



Regulatory risk: Regulations governing the use of alternative data are still in the early stages of maturity. There are open questions about acceptable practices with respect to the use of alternative data

ALGORITHMS

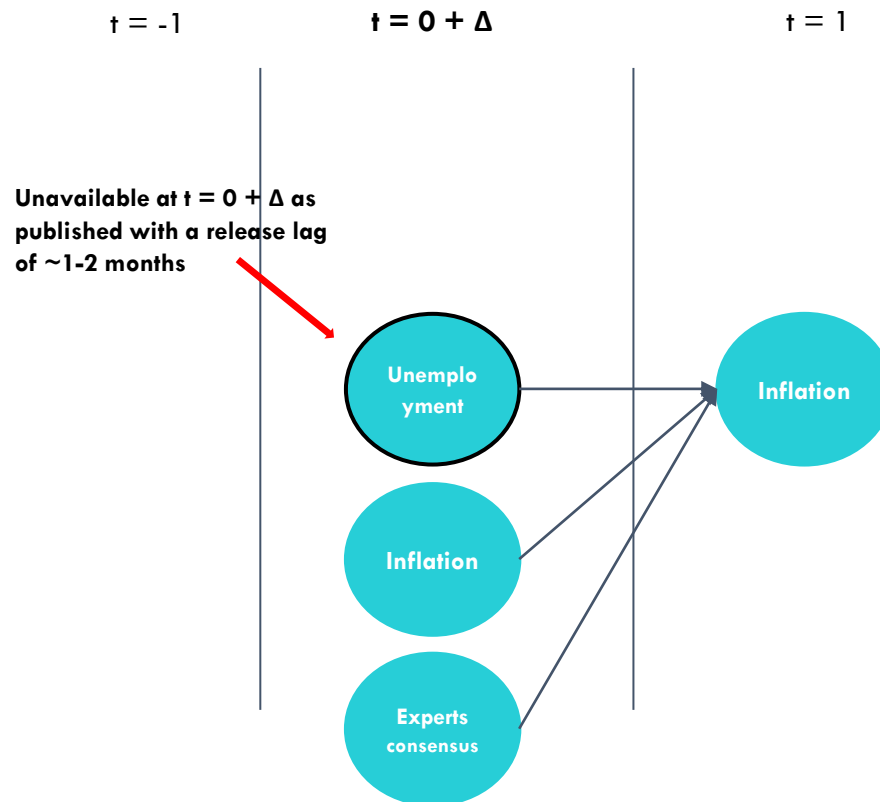
Novel algorithms are needed in alternative data world to deal with the variety of potential explanatory variables as well as their (big) number

- **Features extraction — requires ML algorithms**
 - Image processing (e.g. CNNs)
 - Text/sound processing (e.g. RNNs)
- **Nowcasting/forecasting — since we deal many more variables when we augment our predictive models with alternative data, we must use algorithms that allow sparsity to avoid overfitting**
 - LASSO
 - Ridge
 - Elastic Nets

NOWCASTING & FORECASTING

Alternative Data can be used both for nowcasting and forecasting or a combination thereof

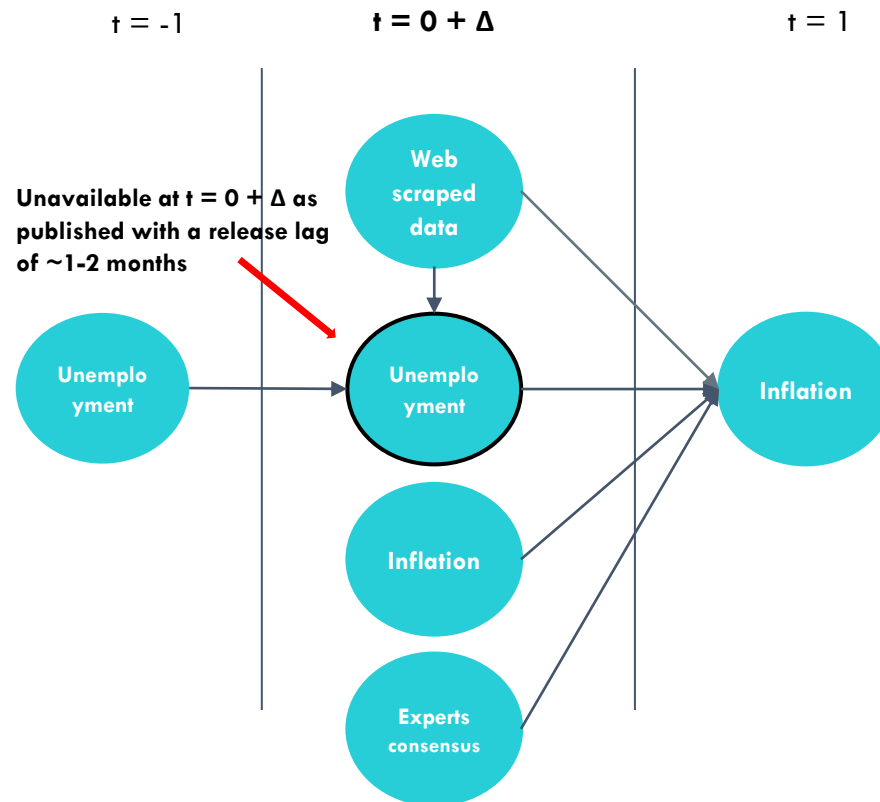
Sometimes our forecasting models can be lacking inputs due to delayed information



NOWCASTING & FORECASTING

Alternative Data can be used both for nowcasting and forecasting or a combination thereof

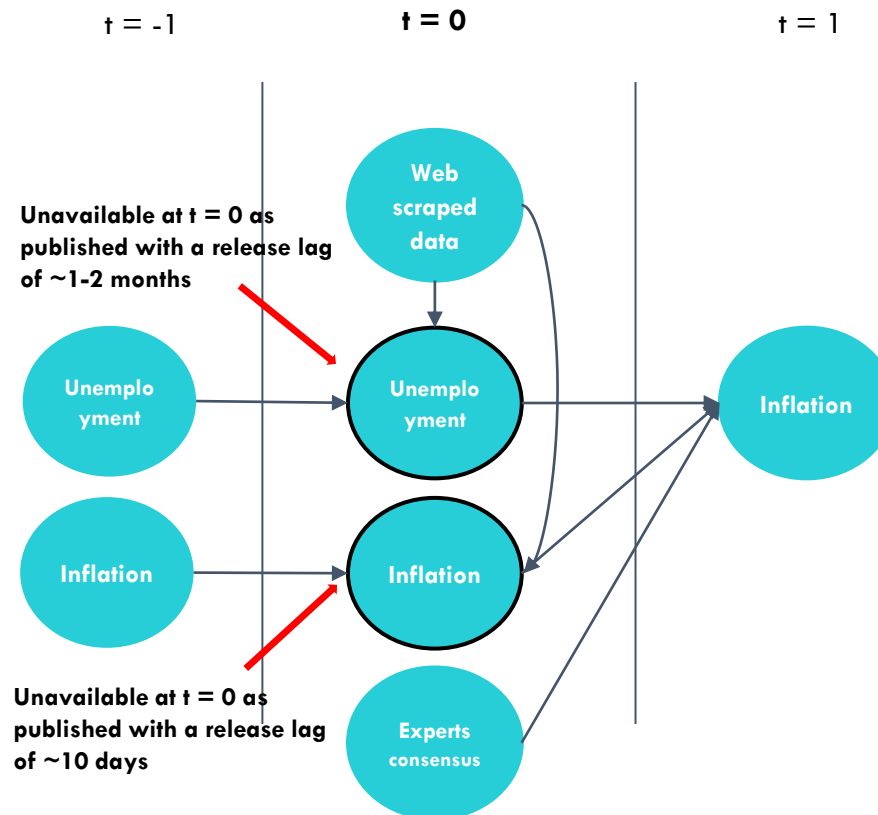
We can build nowcasting models to infer the missing variables



NOWCASTING & FORECASTING

Alternative Data can be used both for nowcasting and forecasting or a combination thereof

We can build nowcasting models to infer the missing variables



SEARCHING FOR ALTERNATIVE DATA



FINDING THE RIGHT DATASET

- **Identify the right dataset**
 - Hypothesis approach: often need to consider what is the question and hypothesis
 - Data driven approach: start with data and then identify the “rationale” for the market tends to be more challenging and easier to have data mining issues
- **Do the analysis to verify the hypothesis**
 - Plotting early on in the process
 - Potentially trying regressions and correlations, with appropriate market data or economic forecasts
- **Clearly, not every alternative dataset will be useful for your purposes**

HOW TO FIND ALTERNATIVE DATA?

- Web directories
 - Can find datasets listed on web (free!)
 - Data scouting & cataloguing firms
 - www.alternativedata.org
- Data firms which aggregate alternative data:
 - Typically take revenue share from underlying supplier
 - Make it simpler to interact with many data firms (one billing etc.)
- Directly to raw data source
 - Corporate firms — but can be challenging
 - Or can collect yourself — time consuming

DATA STRATEGISTS/SCOUTS

- **Within funds, there are data strategists, who**
 - Search for datasets
 - Act as bridge between external data firms, and internal portfolio managers and data scientists
- **External data scouts**
 - External firms in this space to help internal data strategists/scouts
 - Act as intermediary between data firms and data users
 - Paid by data user (i.e. buy side), not by data firms

COSTS OF DATA

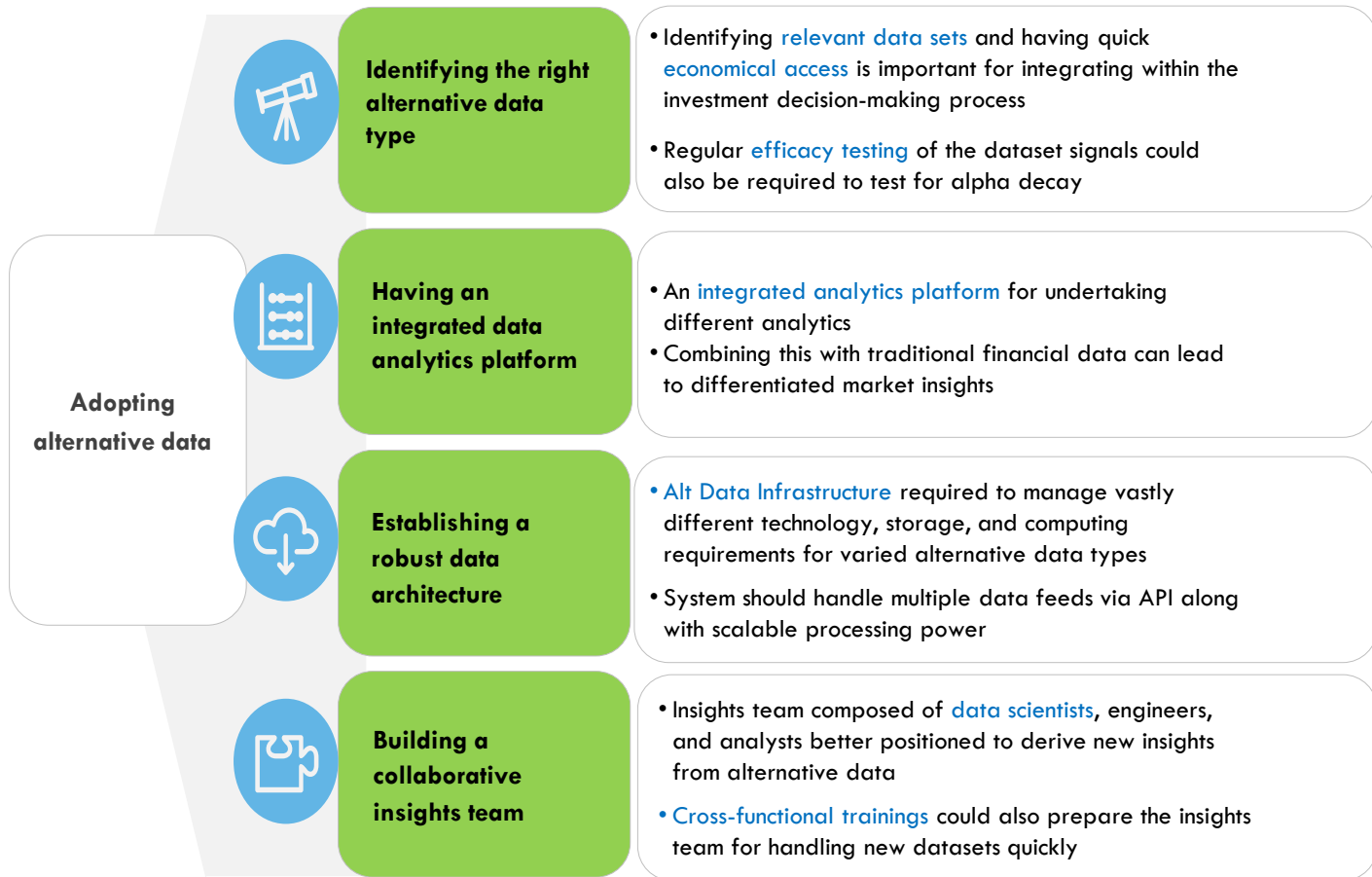
- Depends on several factors
 - Asset coverage
 - Frequency
 - Uniqueness
 - Available trials (mixture of paid/free)
- Can reduce cost by
 - Accessing dataset by company (ie. only those companies you are interested in)
 - Getting lagged data (which is fine for long term investing)
- Most datasets are under 100k USD annually (some can be a lot more, but a rarer)

**EXTRACTING
VALUE FROM
ALTERNATIVE
DATA**



IMPLEMENTING ALT DATA STRATEGIES

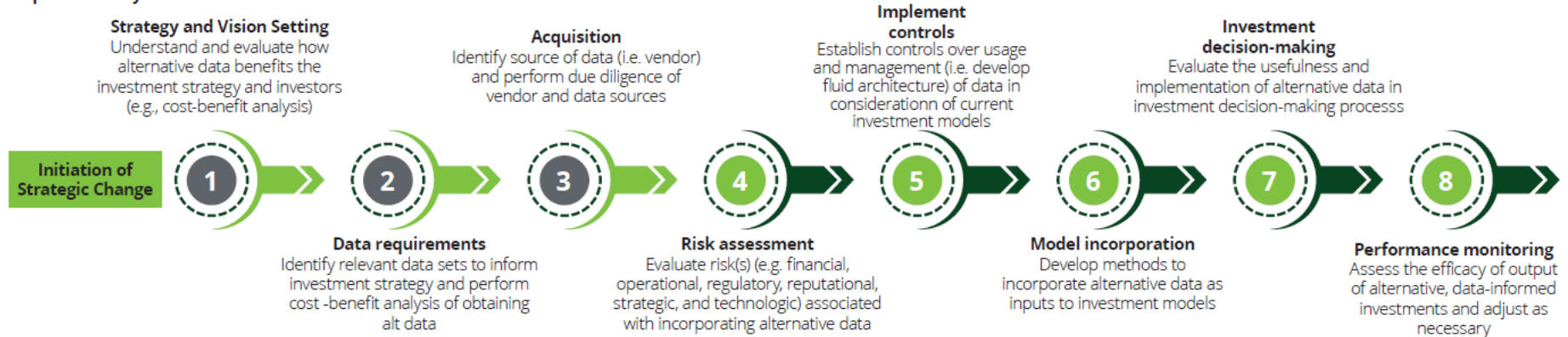
Points to consider while adopting alternative data



THE INFORMATION VALUE CHAIN

In order to realise maximum value for the data assets a combination of prioritization, enhancement and analysis is required, together with a sophisticated valuation structure that reflects the value of data assets to the firm

Sequential lifecycle activities



Thorough risk assessment is required throughout the value chain to ensure that the data stored within the vendor and delivered to customers is regulatory compliant, technologically robust and ethically sound

**CASE STUDY:
GEOSPATIAL
INSIGHTS
SATELLITE DATA
TO ESTIMATE EPS**

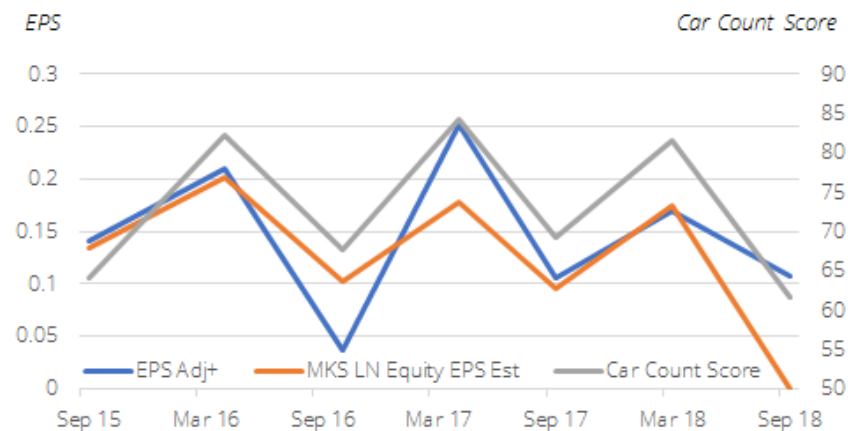


GEOSPATIAL INSIGHTS: RETAIL WATCH

- It is well known that satellite photography can be used to help forecast earning per share for retail stocks
- Has been used extensively in US markets (Orbital Insight), but not as much for European firms
- Uses car counts as a proxy for retail activity
- RetailWatch covers a number of European retailers (both publicly traded and private companies)
- Relatively new dataset

USING CAR COUNTS TO ESTIMATE EPS

- Created a car count score based upon the 6 months of activity related to the earnings period
- Compare against the software company's consensus and actual EPS
- Present results for a multinational retailer:



Preliminary Results from The Book of Alternative Data (Wiley) est 2020