Big Data and Investment Management

The application of data to product development and client satisfaction

BNY MELLON

Table of Contents

Understanding big data is the next challenge and opportunity for the investment management industry

1 Executive Summary

As concepts go, Big Data is one from which it currently seems difficult to escape. The term refers to collections of data sets that are very large and complex, and as such resist attempts to process or manage using traditional methods. More recently, the term has also come to encompass the use of specialised analytical techniques and computing tools to analyse these large and complex data sets. Today's computing tools can quickly interrogate these huge data sets, revealing hitherto untapped trends, patterns and correlations, from which new insights and predictions around current and future purchasing, or in this case investor needs, can be extrapolated.

Volume, variety, veracity, velocity – these are all characteristics of Big Data. The quantity, speed and diversity of information flows continue to expand at a geometric rate, further swelling the pools of available data to be analysed and acted upon.[1]

As the lines between front, middle and back office continue to blur, smarter data management is essential for effective fund management. Big Data facilitates that – but also poses challenges. Through an understanding of these inherent opportunities and potential obstacles, the investment management industry can use their own data to design, manufacture and market solutions more effectively with a view to generating outcomes that are more aligned with investor expectations. The optimisation of portfolio construction may itself offer additional benefits via the mining of data patterns and generating new insights, which in turn may be transformed into additional improvements.

Supermarkets have developed sophisticated data-driven profiling tools. The investment management industry already possesses similar transactional or investment data. Today, 'social data' offers another window into consumer behaviour and bias and should be treated as another valuable additional data pool if it can be harnessed. Identifying and aligning correlations across these two data sets has the potential to generate better and/or more appropriate outcomes for investors.

The paper that follows looks to put history and context around Big Data, examine its credentials and potential as a transformative tool in the current era of shrinking margins and ever-more sophisticated and powerful analytical tools and address the question of how it can be used to solve the complex and commercially-critical issue of enhancing sales performance and client satisfaction.

1 Malvey, Jack; Shrowty, Ashish; Akoner, Lale (2013). BNY Mellon Investment Management: A First Perspective: The Transformational Influence of "Big Data" on the 21st Century Global Financial System

2 Introduction – A Brief History of Data

In August 1854, Dr John Snow, a Yorkshire-born surgeon living and working in London, witnessed the outbreak of a severe cholera epidemic in the city's Soho district. At that time, the accepted view of this disease was that its transmission was through the air – the miasma. Snow had another theory. He claimed the disease entered the human body directly by mouth. Snow explained that the 1854 outbreak was a direct result of sewage present in the water supply and pump in nearby Broad Street:

"…There are no sewers… and the refuse of all kinds consequently saturates into the ground in which the [water] pipes are laid. I found that the water collected by the people, after throwing away the first portion, still contained more organic matter than that supplied to the adjoining streets."[2]

Through his insight and analysis of the patterns of data in the geographical area surrounding the Broad Street pump, Snow saved hundreds of lives when he tiptoed out one night and unscrewed its wooden handle. He went on to save the lives of countless more when he published his theories on cholera to universal acclaim and acceptance. Snow later became known for creating the branch of science now known as epidemiology.

Can we adapt Snow's approach and embrace Big Data to understand the patterns, trends and causes-and-effects in the pools of data that have accumulated all around us? As with Dr Snow, the patterns and correlations in investment data may lead us to conclude that specific changes in how portfolios are designed can drive more effective fund sales linked to improved consumer outcomes.

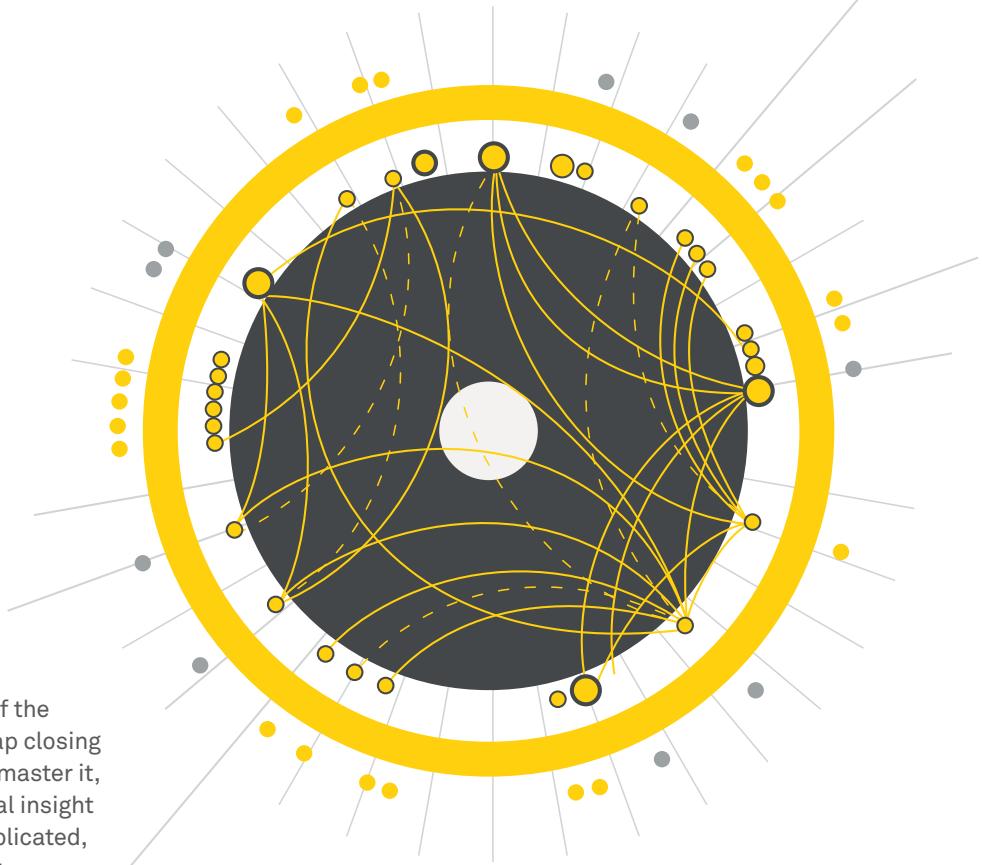
The abridged story of data, Big Data in particular, begins in the design of the database itself. A database is an arrangement of information. It is information architecture. Big Data's visionary forebears are as important as Le Corbusier, van der Rohe and Foster are to the design and development of the built environment. But the technology geeks are far less well known.

In 1970 IBM's[3] E. F. Codd was the first to define the relational model of data that enabled computer scientists to see inside dark pools of data by exploiting tuples. Tuples are ordered mathematical lists of data elements such as rows, columns and tables. Suddenly data was in 3D and could be organised along lines that people could better understand. A very unlikely Prometheus, Codd snatched a new power from the gods and gave it to humans – the ability to organise and make sense out of complex data. Relational databases evolved and grew and became a standard proposition powered further by subsequent developments in the speed and accessibility of computer processing. These new abilities accidentally triggered the routine collection of huge pools of data by corporations and governments, often as unexpected by-products of more mundane daily tasks and compliance requirements to retain information.

In the last decade, databases have suddenly begun to develop fast and reliable new data capabilities such as cubes, business intelligence and super-fast data transfer pipes. Now there is Apache Hadoop and Google's MapReduce, programme frameworks and algorithms which take the question to the data and hold entire databases in dynamic, slender strings like twisting spaghetti on a fork. Looking inside data gets faster and easier. So, why use a small sample data set when you can interrogate all the data that ever was, right now? This relentless process of data evolution has enabled the machine to nose just in front of us. Databases now provide high quality analysis using fast, reliable queries across multiple dimensions as a starting point.

This explosion of machine power and speed means we humans are catching up in our understanding of the potential of data. Overall, we have scaled up our own understanding and application of data science

2 Snow, John MD (London, 1849). "On the Mode of Communication of Cholera"
3 Codd, E.F. (1970). "A Relational Model of Data for Large Shared Data Banks"



to cope with the new size and usefulness of the data being collected around us. With the gap closing between the size of data and our ability to master it, we have only now just begun to produce real insight using correlations across big pools of complicated, often unmatched, data sets.[4] By preventing overlearning – drowning intelligent machines in too much data – human guidance now controls the analysis of data again. The stage is set for Big Data itself. And as before, it's down to the quality of the questions we ask.

It has taken a series of specific breakthroughs in computer science and a corresponding development in our ability to manipulate data with ease and speed to enable us to now have the power of insight into the increasing amount of data that is being captured. We now have the ability to peer into the huge pools of information which have ballooned in size around us and use this information in a meaningful way.

Big Data is crystallising into a method used by engineers and data scientists to investigate and analyse information by applying inquisitive analyses and correlations that can be accurately deduced from staggeringly huge sources of data available – even live data. But it has taken human engagement to apply these powerful new tools to effect change.

In 1970, IBM's E.F. Codd defined the relational model of data by exploiting tuples.

4 Mayer-Schonberger, V, Cukier, K (2013). "Big Data – A Revolution That Will Transform How We Live, Work and Think"

3 Data Sets

The era of Big Data is now upon us. Full-page IBM advertorials in the media broadcast news of a revolution in predictive analysis and more. On opposite pages are written portentous articles in dedicated column inches of big data "news". Wikileaks and Edward Snowden have become rock-star whistle-blowers, the sky is falling on our heads but consumers and industry react with apathy. Despite the outrage, the news just seems to reinforce the awkward modern paradigm in which we find ourselves. In a world of Big Data, we submit ourselves to the systematic, thorough-going collection and storage of all our personal, emotional and financial data online. We sometimes worry that there are benefits to others in this digital theft – for example, supermarket rewards points. And so we half-heartedly resign ourselves to the knowledge that Big Data is the online equivalent of someone going through our rubbish bins. But why should we really care? We have nothing special to hide. There must be more.

When John D. Rockefeller started Standard Oil in 1870, refineries in the US processed crude oil primarily to recover the kerosene. There was no market for the more volatile fractions, including gasoline, which were considered waste and were often dumped directly into the nearest river. It was only after the invention of the automobile that demand shifted to gasoline and diesel, which remain the primary refined products today. We are in a similar place with Big Data and the investment management industry. Today, all types of data points are thrown away or ignored. As with the early kerosene refineries the waste data is discarded as there are no products in the marketplace to be fuelled by this data and no apparent need to refine it further. But there is hidden value. The data contains potential value for the regulator, the manufacturer and the consumer. Of particular value are the governance, control, specification and appropriateness of the data in respect to investment management performance and this goes right to the heart of the opportunity for the investment management industry.

Custodians, depositary banks and administrators are actually at the forefront of product development in Big Data solutions. These industrial scale functions have direct proximity to, and operational expertise of, the requirements and products of investment managers. In the modern investment management industry, there are complex requirements to maintain fund, settlement, reconciliation and portfolio position data, all of which are reconciled, audited, processed for exceptions and stored for historical and analytical purposes. Examples of the types of data held by custodians and asset servicers are:

CUSTODY – the position keeping, identification and storage of detailed investment and transaction data across hundreds of thousands of investment portfolios, is held as passive data. It is a store of asset ownership information connecting the instrument held to its nearby or 'proximate' investment manager or discretionary manager and identifying the ultimate beneficial interest.

ADMINISTRATION – transaction, settlement and cash reconciliation position data alongside an integrated and growing set of geographical and other 'non-financial' data for example white-label and outsourced online consumer journeys that the administrators increasingly integrate with.

> In 1870, John D. Rockefeller started Standard Oil to produce kerosene – gasoline was dumped because the automobile hadn't been invented. The investment management industry is doing the same thing with Big Data.

Big Data and predictive analysis can enable investment managers to see 'inside' their clients' portfolios and underlying investments. Here is a real event as a test case, a starting point to examine the case for more efficient distribution and better investor outcomes. Auto-enrolment (AE) in the United Kingdom is a new programme of mass compulsion for workplace saving, as set out in legislation under the UK's Pensions Act 2008. The AE programme is a sea change in the UK investment management industry that is now underway and will last until 2017. It will enrol 9 million new savers into a national savings programme and is arguably the single biggest change in the UK investment management industry since the end of the Second World War. The goal of the legislation is to oblige citizens to provide for their own retirement incomes by investing throughout their working lives. However, because of its unique circumstances AE will also create a huge new digital footprint in the investment management industry itself. The AE data collected will cover the full spectrum of custodian, asset servicing and third party administrator (TPA) data alongside emotional, sentiment and social data collected in respect of investors themselves. These combined data sets give the investment management industry a major and rare opportunity to analyse, design and potentially manage investment portfolios differently.

AE's big data is raw material from several key sources:

- pension scheme and individual policy books and records;
- settlement, trading and transaction data from central securities depositories and custodians;
- anonymised bank account transaction records;
- instrument, securities and asset data such as corporate actions information;
- data regarding the relationships between tax, government and regulatory returns and employer-held information, such as contribution records;
- social network analysis data, such as Tweets and Facebook "Likes", email keywords and cultural feedback loops.

4 Dark Pools of Data, Predictive Analysis and Behavioural Finance

Outlined in the following section is an experiment that shows an example of how investment managers could seek to utilise Big Data; how patterns and insight of data held by custodians, asset servicers and TPAs could facilitate sales and improve investment outcomes in the context of AE. These are the four stages:

1. **Dark Pools of Data: the reservoirs of investment management and portfolio information which are the raw materials of analysis into the correlations and connections of AE;**
2. **Predictive Analysis: the activity of human and machine interpretation of that data which analyses the patterns and trends found inside it;**
3. **Behavioural Finance: the method of psychological interpretation used by wealth and portfolio managers to determine consumers' attitudes to risk and the proposed corrections required in portfolio asset allocation to compensate ('correct') for that; and**
4. **Creating an Outcome: the process of applying analysis of the investment industry data to improve product design, drive sales and potentially generate improved investor outcomes.**

Dark Pools of Data

The experiment contemplates the potential correlations, connections, patterns and trends within dark pools of investment management industry data. These pools may be isolated or integrated and are big, complicated and often apparently useless stores of raw data. At face value, the dark pools are in decline; incapable of being applied to any structured enquiry or at all. But this is the raw material of the experiment. Ordering the data into a structure is not necessarily a prerequisite. What is required is the creation and editing of, a sort of high-level map or key to the dark pools of data. This 'data dictionary' is a human artefact that helps navigate the data – it contains meaningful explanations of all the fields and rows in the data, in ordinary language. Together with examples in 'plain speak' the data dictionary is the link between people who are subject matter experts and the technical data scientists. The data dictionary brings order to chaos. It is a first step towards insight into the investment management data and is an important asset in itself.

Where can we find these pools of data?

1) **Investment management industry portfolio databases**, the subject of an Eagle Investment Systems (Eagle) white paper[5], store investment instrument reference data including security identifiers, pricing data, analytical data and investment portfolio data. The Eagle paper sets out that the unconnected silos of investment management data may be better managed by their owners through a continuing programme of data quality, enrichment and special considerations (such as derivatives exposure values and returns, including cash values). The Eagle paper sets out the challenges of capturing and managing such dark pools of investment management industry data, including heterogeneity, incompleteness, scale, 'dark silicon', privacy, timeliness and the assistance of advanced programming techniques such as Google's MapReduce and Apache Hadoop. Eagle proposes that this is a task worth executing for investment portfolio managers, wealth managers and fund managers.
2) **Custodian banks, asset servicers and TPAs** are data controllers and processors; storage hubs that connect the underlying investments of retail and institutional portfolios. Consolidating the custodian and TPA data is an impossible task without the use of a comprehensive data dictionary. A new wave of consolidation in the custodian, asset servicing and TPA industry means ever bigger data pools. The industry has responded with the evolution of multi-function solutions, such as international central securities depositories (ICSD). Custodians, TPAs and ICSDs can structure and consolidate their expanding dark pools and begin to offer their clients regulatory, marketing and advertising insights from product development and design. Is there a market for intelligence on data that was formerly thought of as waste?

5 Jenkins, T (2011). "Investment Portfolio Data Management"

3) **Transactional data** is a huge resource that is growing exponentially and provides a complex and reliable set, ripe for analysis. A good example is how an innovation lab for SAP technologies, created a prototype transactional analysis system on behalf of a bank for non-expert users. The device enabled users to filter for a subset of bank card transactions by entering parameters for all card holder's data such as a selected age, gender, marital status and postcode. The resulting graphical display on a touch-screen listed bank card transactions carried out by that subset in real-time and by product and supplier (for example, breakfast cereal at a local supermarket in total monetary value).

4) **Personal data:** We express ourselves socially on the internet every day. Through personal blogs, Instagram, Twitter, Facebook, LinkedIn and Google as Search, Gmail, Google Docs, Google Calendar, Google+, Hangouts, Places, Latitude, Android, Chrome and text messages. We produce an ever-expanding personal digital exhaust trail across the internet and as we do so we are literally creating our own unlimited dark pools of data. It was these pools of data that have been the subject of the Snowden / NSA and newspaper revelations and are the focus of the 'metadata' debate in national and international politics.

5) **Community data** is the wriggly, unpredictable and most fascinating collective dark pools of data that groups of people, wittingly or otherwise, create. It is emotional output online. Sentiment taken raw from all of us. Smile for London is a community art and culture project which in 2012 launched **Word in Motion**[6], an exhibition that blended the world of literature with the world of art with an audience of 1.5 million commuters and included poetry and prose by Jarvis Cocker, Ray Davies and Cerys Matthews. App-based responses captured emotional Big Data about the community in a large and highly disparate data set. At the other extreme, DataSift enables marketers to sample 'live' emotional feedback from global consumer bases, in real time. By looking for keywords and phrases in internet channels such as Facebook, Instagram, Twitter, Google+, YouTube, blogs, Amazon.com, Reddit, Flickr, Wikipedia and many more, campaign managers can taste, capture and live the emotional state of the online planet.

Predictive Analysis

There are dark pools of investment management data, sentiment and emotional data out there. With so little structure, how can we possibly see trends and patterns? Like the infinite variations in the Gulf Stream, surely only the most obvious patterns are visible or even useful? Are we in danger of drowning in data, a blizzard of information that we can't utilise?

Big Data is human and machine analysis backed up by a set of mechanical processes. There is a requirement for machines and data to link together initially, enabling humans to think dynamically and analyse the patterns in the incoherent and disparate pools of data. There are hidden and sometimes seemingly random similarities that only analytics can identify. Through machine learning, correlations and patterns can be identified that cannot be observed manually or even with traditional analytical techniques that require significant human guidance. However, people can interpret correlations and distinguish between patterns and meaningful relationships. Based on this insight, they can give further guidance to the analytical tools to push further and dig deeper.

That leap is the extra value in Big Data. It is the very human work of predictive analysis; when we look at the patterns and say – what if? What if we took that pattern, made an assertion about it and asked another question about it? What if we married the pattern with for example, behavioural bias compensation? It is our human interaction with investment management industry data that makes all the difference between observation (of patterns) and insights in understanding.

Next we consider some of the habits learned and milestones achieved by examples of Big Data in action using an example from a financial institution counterbalanced by another example from an entertainment distributor:

a. Driving by Looking in the Rear View Mirror?

In 2005, Chase Manhattan[7] published an internal paper on the power of financial product data analysis. The paper uses predictive analysis across Chase's retail financial consumer data sets, identifying its own consumers most likely to refinance mortgages and thereby enabling the bank to intercede and propose an attractive mortgage interest rate to them. By lowering churn rates on profitable customers, the Chase predictive

6 Kotas, J, Stevens, A (2012). "Word in Motion 2012"

7 Moazami, A, Li, S (2005). "Mortgage Business Transformation Program: Using CART-based Joint Risk Modeling (and including a practical discussion of TreeNet)"

programme also identified those consumers not suitable for mortgage product offers from the bank. This category were remaindered into collateralised pools of mortgage obligations and sold on for commercial securitisation with other banks and finance organisations.

Despite the part played by mortgage-backed securities in the global financial crisis, it was the value to Chase in knowing which consumers to target with which product that offered a commercial advantage. The bank used Big Data as the means by which the target group of consumers could be identified and action taken. Insight is at the core of predictive analytics.[8] Simply looking at the numbers can provide an observation from which to proceed but it requires true insight to know what to do with that. For this we need people.

b. The Power of Communities

In September 2009, the Netflix Prize 1$ million Grand Prize was awarded to team "BellKor's Pragmatic Chaos" for their work using viewer preferences to enhance the accuracy of recommendations for other films the viewer might also appreciate.[9,10,11]

Faced with the obstacles of innovation and the constraints of an incumbent data analysis provider, Netflix decided to offer a cash prize to the community of movie-loving data analysts to deliver a better predictive algorithm for audience appreciation of upcoming movies. The community delivered – outperforming the existing predictive reporting by a factor of 10% or more. This confirmed a long-held suspicion: that innovation and analysis of big data can best be done in the wild – not just data capture but analysis. Many hands lighten the work.

Behavioural Finance

A last ingredient in our working example of the application of Big Data to the investment management industry is the science of behavioural finance. Is behavioural finance a fashionable insight into the complexities and performance of the wealth and investment industry or voodoo in the unpredictable and risky world of financial advice? Is this a trend or a paradigm shift?

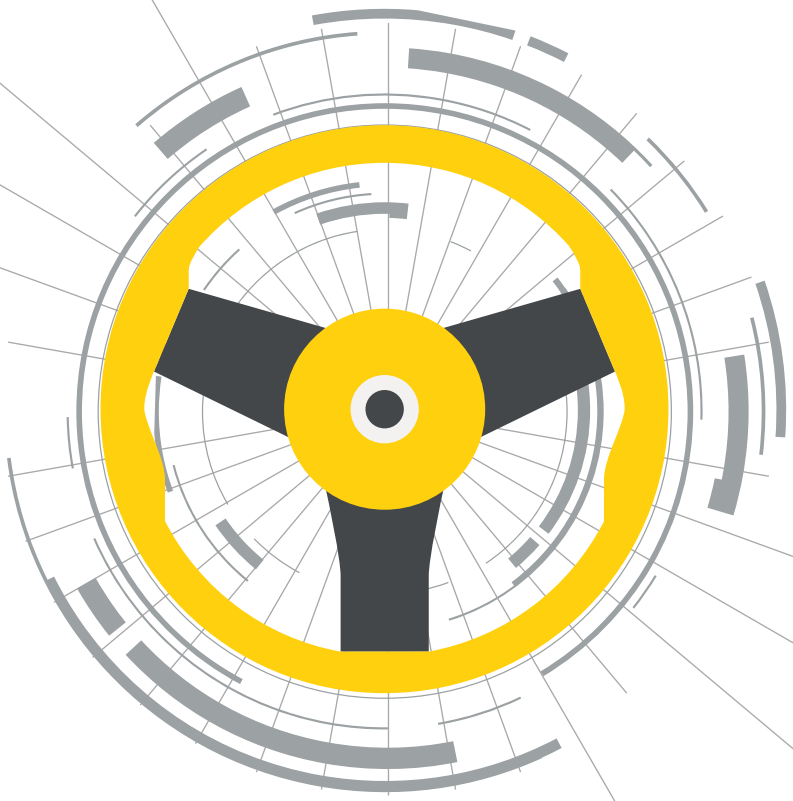
Delving into the minds of investors can reveal correlations between human personality types and their preferences and concerns when investing. Michael Pompian[12] has researched and advanced the science of human attitudes towards investing and concluded that psychology and the study of the human being can deliver better understanding of investment management aversions, desires and drivers. By categorising people into approximate types, Pompian postulates that each type should have a corresponding investment management portfolio suited to temperament, loss aversion, risk appetite and so forth.

Davies & Haisley[13] have extended the concept of behavioural finance further to include a utility calculation that measures an individual's emotional personality in respect to his/her anxieties about financial losses versus their desire to make financial gains. A sort of 'happiness' index. Applied to the investment management industry, this approach attempts to place investors in the correct investment portfolio for their specific 'risk-return' appetite. Measured by outcomes, the utility position of an individual human investor means the chance of a more refined, better-attuned investment management portfolio for each investor and an expectation of improved outcomes.

An extension to the behavioural finance method – via Big Data – might not be obviously viable or beneficial. Applying data of the 'live' emotions of a group of individuals to the investment management industry, for example newly-auto-enrolled co-workers and a level of exactitude and relevance to the underlying behavioural status of the group would be extremely challenging.

However, taking the basic algorithms of behavioural finance and coupling them with the messy correlations of predictive analytics and the scale of dark pools of investment management industry data ought to reveal real observations that can be translated into valuable insights about mass investor sentiment and their best-aligned risk portfolio.

8 Kalakota, R & Netke, S (2012). "Predictive Analytics 101"
9 Koren, Y (2009). "The BellKor Solution to the Netflix Grand Prize"
10 A. Töscher, M. Jahrer, R. Bell (2009). "The Big Chaos Solution to the Netflix Grand Prize"
11 M. Piotte, M. Chabbert (2009). "The Pragmatic Theory Solution to the Netflix Grand Prize"
12 Pompian, Michael M (2012). "Behavioural Finance and Investor Types"
13 Davies, G. B., Haisley, E (2013). "Overcoming the cost of being human (or, The pursuit of anxiety-adjusted returns"



Analysing Big Data produced by the investment management industry will provide observations from which to proceed. But it requires true human insight to know how to use them.

5 A Thought Experiment

The theoretical experiment for dynamic investment portfolios runs as follows:

What if, over the next three or four years, the combined workforces of small to medium-sized enterprises (SMEs) in the UK, adopt AE and cloud-based mail systems in parallel? In other words by 2017, 9 million new savers are also the same cohort of 9 million new users of company Gmail.

In this fictional world, an investment management industry provider (such as a custodian, asset servicer or TPA) asks these SME employees to 'opt-in' to share their anonymised company email content to the provider's Gmail app. This would not be any part of their employer's Gmail app, this is an important point.

The resulting data for a SME workforce, for example a 300-strong widget manufacturer in Bedfordshire (UK), is then captured by the investment management industry provider and predictively analysed using a behavioural bias algorithm provided by an investment manager. Working out the 6-monthly or annual emotional feedback of the total workforce will enable the investment managers to correlate the predictions against the investment management industry provider's big data and apply the correct investment portfolio for each of the cohorts in the workforce, corrected for bias.

Previously this accurate and predictive investment management service was only available to high-net worth individuals and even then, only on an ad hoc basis. Whilst being aware of the dangers of too-frequent dealing or rebalancing (where the costs and inconvenience of transactions often outweigh the benefit of the rebalancing) this evolution should improve the satisfaction and comfort of the workforce, encourage better engagement with the workplace savings habit and potentially results in a more suitable investment outcome before retirement. This is an excellent and obvious extension of the services and products of the investment management industry which implicitly drives sales and efficiency.

6 Creating Better Outcomes for Managers and Investors

Our working example into the application of investment management industry data to consumer portfolios (in this case for pension AE) concludes with an attempt to bring it all together.

What are the drivers and other factors away from the core investment industry data used and collected by custodians, asset servicers and TPAs? What are the links between the investment industry data and how the moving parts of the experiment itself operate? The following summary attempts to set these out:

- **WHY THE WORKPLACE?**
  We have now entered the era of the individual investor. The mass market consumer experiment that is AE in the UK has commenced. This legislative change, based on its Australian and US forerunners (Superannuation and 401K, respectively) enjoins a new 9 million consumers with an obligation to contribute directly from their payroll and into a workplace savings scheme.
- **HOW DOES BIG DATA FIT IN?**
  The cohort of new auto-enrolled workers in the UK is historically atypical. Crucially these new investors will produce their own (new) investment management industry data and can be isolated within data sets. Additionally, the new investors will also reveal their attitudes, psychological personality types and biases which have not been captured or analysed previously. Transactional, personal and community/emotional data will be shared by this cohort as they journey through their investment lifecycle.
- **CLOUD MAIL ADOPTION BY SMES.**
  This is happening in parallel. SMEs are adopting cloud mail for workers and colleagues over the same timescale as firms are staging for AE. Cloud mail means the end of the server rooms for SMEs and the welcome advent for employees to bring their own devices. Secure cloud-based emails and storage of documents means a new and timely pool of data for the investment management industry's big data experiment.
- **ARE THERE ALTERNATIVES?**
  If it were the sole preserve of Google, SAP or Microsoft, then we would already have Big Data solutions for the investment management industry. But data engineers and scientists work largely in isolation, or at least on a single task thread. It is a business driver in each case which demands and then pushes forward commercial solutions. Those lessons about product design through experiment and feedback loops – especially emotional data – are crucial for the next steps of data evolution.
- **A DAY IN THE LIFE OF A DYNAMIC INVESTMENT PORTFOLIO.**
  Big Data is a tsunami of information, in this case the raw data from the investment management industry. The data is unstructured, complicated, volatile and incoherent. Wrangling the data and transporting it between data stores is a mechanical task. But it is the data engineering which is the interface, the coalface, to the wider public via the algorithms and science of behavioural finance.

Fusing the investment management industry's dark pools of data to the emotional data of a workforce by means of behavioural finance and risk-profiled portfolio selection tools that would create an always-aligned risk profile and better placed investors is a testable outcome. Financial personality analysis tests match an investor's attitudes to loss and gains. These are typically carried out infrequently, sometimes only once, at the outset of a portfolio. Coupling all the dark pools of data set out in the investment management industry alongside consumer data (such as email and sentiment data) reveals clues about the correct portfolio for groups of investors.

In order to make this 'dynamic' involves analysing live transactional data, predicting what it means for the cohorts and segments of consumers involved, considering this against the patterns, trends and correlations within the investment management industry dark pools and finally extrapolating a range of outcomes for investors.

7 Conclusion

Jane Austen said of wealth

"...It is a fact universally acknowledged that a single man in possession of a good fortune must be in want of a wife."[14]

But Austen's characters fished in a small gene pool. How much misery and heartache would the characters be spared if Jane Austen lived in the world of internet dating?

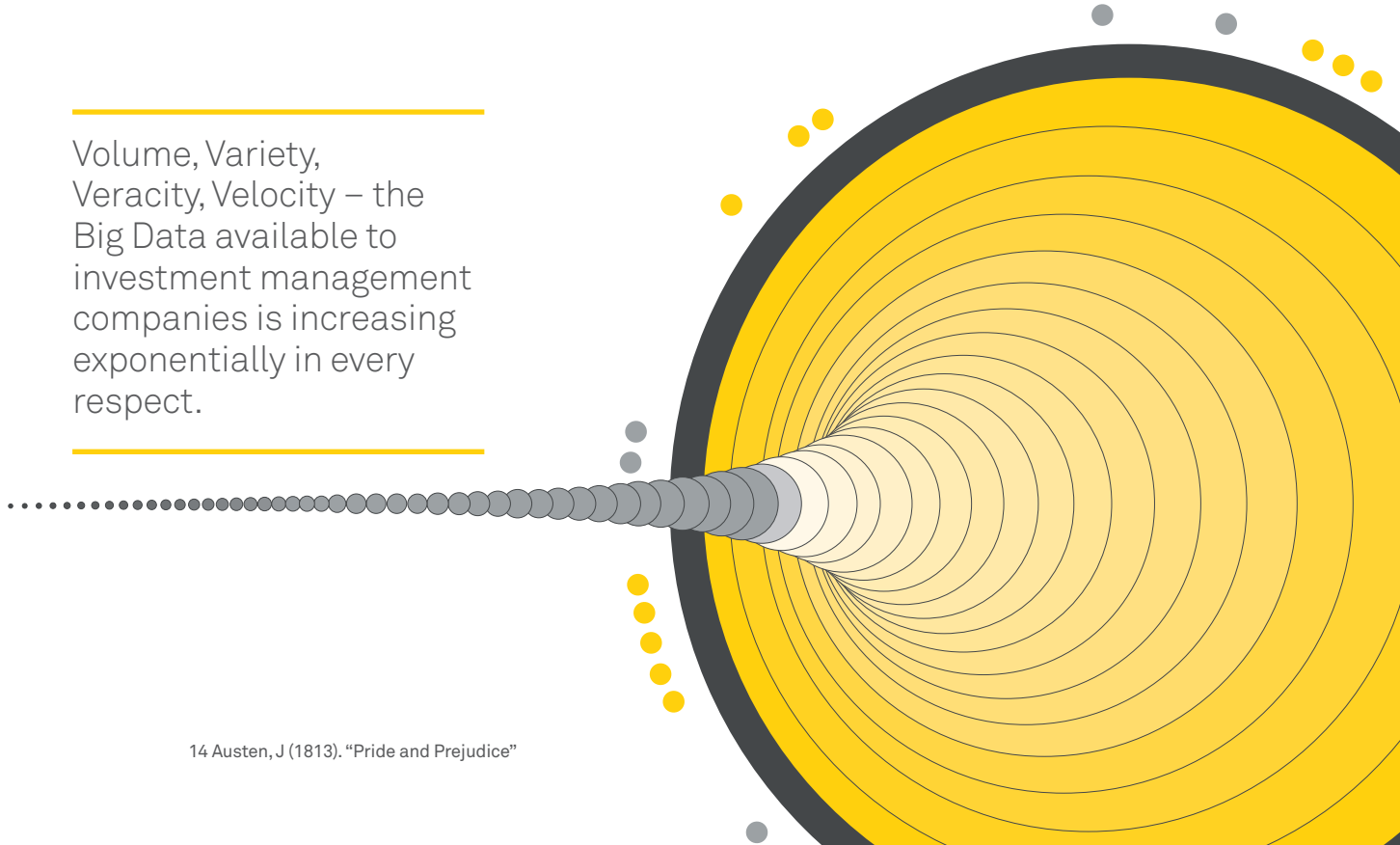
It seems an obvious challenge but 'big data, so what?' So what if there are petabytes of data collected daily about the weather? Who cares how many train tickets are bought online in France? Why should Chinese search words on Google be of interest in an outbreak of avian flu? Why do air traffic patterns matter for the environment and fuel efficiency?

Human curiosity never sleeps. John Snow's is not the only compelling story. It is relevant – when we hold and analyse data we see shapes and patterns, the way we have evolved to do so from the Stone Age. More data must equate to more patterns and better understanding.

Big data has a bewildering plethora of sources, structures and storage silos. Its predictive algorithms are a parallel and complementary art; there are a handful of existing Big Data hubs or transports to share the databases and fewer still humans capable of understanding or driving real predictive analysis across whatever network. But by accepting messy data's flaws, we are able to exploit its inherent correlations to our advantage. We have proposed this in the greenfield of AE data, mapped to behavioural finance, as a big data challenge that is both academically verifiable and commercially relevant. John Snow would surely have approved.

Finally, and not simply as a coda, is the relevance of the investment management industry itself. Our industry has long been like Rockefeller's Standard Oil – generating and storing the raw material of data but refining only a small portion of that for use and exploitation. Now, in the age of powerful data analysis tools and shrinking margins, it is the investment management industry itself that has an opportunity to innovate and widen the scope of data-based solutions. By the mechanism of experiment and case study, this can be focussed and commercially advantageous both for the investment manager and the investor.

Volume, Variety, Veracity, Velocity – the Big Data available to investment management companies is increasing exponentially in every respect.



14 Austen, J (1813). "Pride and Prejudice"

8 Bibliography

Malvey, Jack; Shrowty, Ashish; Akoner, Lale (2013). BNY Mellon Investment Management: A First Perspective: The Transformational Influence of "Big Data" on the 21st Century Global Financial System

Snow, John MD (London, 1849). *"On the Mode of Communication of Cholera"*

Codd, E.F. (1970). *"A Relational Model of Data for Large Shared Data Banks"*

Mayer-Schonberger, V, Cukier, K (2013). *"Big Data – A Revolution That Will Transform How We Live, Work and Think"*

Jenkins, T (2011). *"Investment Portfolio Data Management"*

Kotas, J, Stevens, A (2012). *"Word in Motion 2012"*

Moazami, A, Li, S (2005). *"Mortgage Business Transformation Program: Using CART-based Joint Risk Modeling (and including a practical discussion of TreeNet)"*

Kalakota, R & Netke, S (2012). *"Predictive Analytics 101"*

Koren, Y (2009). *"The BellKor Solution to the Netflix Grand Prize"*

A. Töscher, M, Jahrer, R. Bell (2009). *"The Big Chaos Solution to the Netflix Grand Prize"*

M. Piotte, M. Chabbert (2009). *"The Pragmatic Theory Solution to the Netflix Grand Prize"*

Pompian, Michael M (2012). *"Behavioural Finance and Investor Types"*

Davies, G. B., Haisley, E (2013). *"Overcoming the cost of being human (or, The pursuit of anxiety-adjusted returns)"*

Austen, J (1813). "Pride and Prejudice", Penguin Books

05/2015 T2127

BNY MELLON