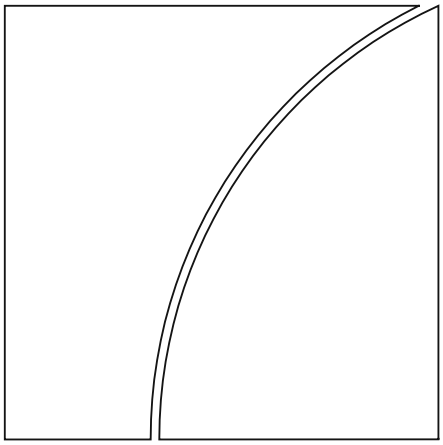
Irving Fisher Committee on Central Bank Statistics



IFC Working Papers

No 14

Big data: The hunt for timely insights and decision certainty

Central banking reflections on the use of big data for policy purposes

by Per Nymand-Andersen

February 2016

BANK FOR INTERNATIONAL SETTLEMENTS

IFC Working Papers are written by the staff of member institutions of the Irving Fisher Committee on Central Bank Statistics, and from time to time by, or in cooperation with, economists and statisticians from other institutions. The views expressed in them are those of their authors and not necessarily the views of the IFC, its member institutions or the Bank for International Settlements.

This publication is available on the BIS website (www.bis.org).

© *Bank for International Settlements 2016. All rights reserved. Brief excerpts may be reproduced or translated provided the source is stated.*

ISSN 1991-7511 (online)
ISBN 978-92-9197-317-0 (online)

Big data: The hunt for timely insights and decision certainty

Central banking reflections on the use of big data for policy purposes

Per Nymand-Andersen[1]

"Progress lies not in enhancing what is, but in advancing towards what will be"
(Khalin Gibran).

Abstract

A new data paradigm has emerged. Despite the human instinct to reject what cannot be fully comprehended, the big data industry is extracting new causations among multiple pools of micro-data that previously looked unrelated. This is leading to new, timely indicators and insights, and may generate new economic theories. Central banks do not have to be ahead of the curve, but they should not miss this opportunity to extract economic signals in almost real time, learn from the new methodologies, enhance their economic forecasts and obtain more precise and timely evaluations of the impact of their policies. Moreover, they should encourage these new data sources to be transparent regarding their methodology, quality and aggregation methods for publishing new types of economic indicators. Lastly, the big data industry will challenge not only traditional statistics and economics, but also the way in which these are fed into the decision-making process. This paper argues in favour of developing a conceptual framework and road map for central banks using relevant pilot studies. The objective is to explore the conditions for making systematic use of these sources as part of the central banking policy toolkit.

Keywords: Big data, statistics, economics, nowcasting, indicators, central banking policies.

1. A revolution in thinking and practice

Over the past decade, big data have become an increasingly important aspect of our daily lives: the term is being used in several scientific fields, in new business models

1 Adviser at the European Central Bank, e-mail contact: per.nymand@ecb.int. The views expressed are those of the author and do not necessarily reflect those of the European Central Bank. The author acknowledges the useful comments made by Bruno Tissot (Bank for International Settlements), Timur Hülagü (Central Bank of the Republic of Turkey) and the support of Heikki Koivupalo (European Central Bank).

for establishing corporates, in governmental discussions and new government policies. Big data have been identified as providing a new service with high growth potential, generated by the continuously changing way in which we live, communicate, socialise, interact, obtain intelligence and exchange information, and by the way in which public authorities structure, operate and interact with the private sector. It is our new digital footprint, logging and combining records of individual actions and digital prints.

Central banks may find it hard to dismiss big data as "fog" – a popular buzz word – that will disappear of its own accord. Big data represent an ever-changing product, one with its own prevailing technical dynamics – a continuously expanding revolution, which affects and ultimately changes the social and economic behaviour of business enterprises, governments and ordinary people. Big data can be defined as a source of information and intelligence resulting from the recording of operations or from the combination of such records. There are many examples of recorded operations of this kind, such as the records of supermarket purchases,[2] robot and sensor information in production processes, road tolls, trains, ships, mobile tracking devices, telephone operators, satellite sensors, images, and behaviour, event and opinion-driven records from search engines, including information from the social media (Twitter, blogs, telephone text messages, Facebook,[3] LinkedIn) as well as from internet information scraping and speech recognition tools. The list seems endless, with more and more information becoming public and digital as a result, for example, of the use of credit and debit payments, trading and settlement platforms, and housing, health, education and work-related records. Annex 2 gives a few examples of the diverse current commercial use of big data, although the list is far from exhaustive.

"Big data" seem to be associated with the ability to combine recorded information and extract intelligence from multiple sources. But the literature[4] provides little evidence of how to define or describe the term "big data" more precisely. What volume of data is needed before the classification "big" can be used and what characteristics are required of the dataset? There is no clear answer as we are dealing with a moving target. "Big data" of ten years ago no longer seem "big" today: every day volumes seem to expand, velocity is increasing and the variety of data sources and formats is proliferating. While Gartner's 3V model[5] seems to have acquired certain popularity, IBM has produced an infographic[6] that provides an overview of the components of "big data" by adding a fourth "V".

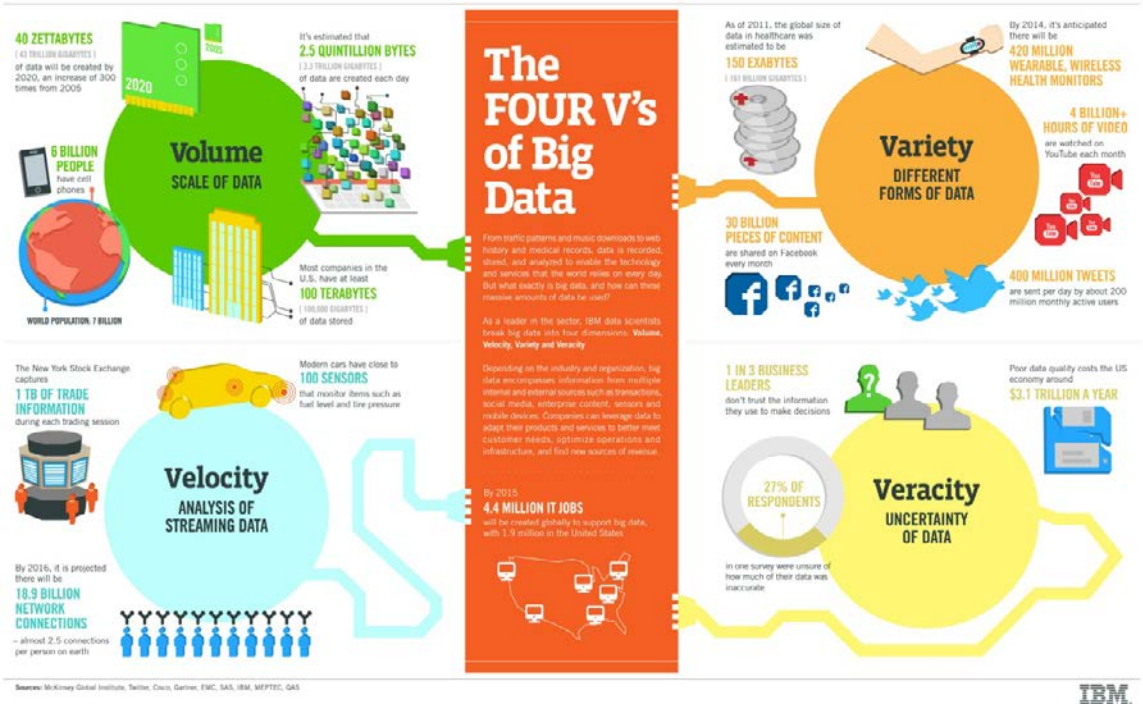
2 For instance, Walmart, a retail giant, handles more than one million customer transactions every hour, feeding databases estimated to hold more than 30 petabytes. One petabyte of digital music would play for 2,000 years.

3 Facebook, the social networking website, contains 250 billion photographs and is growing with 350 million photos uploaded every day. http://www.theverge.com/2013/2/22/4016752/facebook-cold-storage-old-photos-prineville-data-center.

4 For information on how big data is defined in the literature, see Annex 1.

5 Laney, Douglas, 3D Data Management: Controlling Data Volume, Velocity, and Variety, META Group (now Gartner), 2001.

6 http://www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-big-data.jpg



Volume; In fact, the term "big data" is so mystifying that some people are now using metaphors to illustrate just how big "big" has to be – in terms of stockpiles of CDs, papers, books stretching from the earth to the moon or the number of days it takes now to create as much information as was available throughout the whole of history and up to a given year.[7,8] Industry has labelled the increase in data as an "industrial revolution"[9] and the concepts of "data" and "information" are actually becoming more interchangeable and increasingly difficult to separate. This is why *"the science of statistics"*, as the means of transforming "data" into "*relevant information*", will have a crucial role to play.

Variety; "Big" is not just about volume: it is also about complexity and multidimensionality. Variety, one of the criteria in the Gartner model, relates to the mix of data sources and nature of the data. In terms of form, data could have metric, ordinal or nominal value, such as transaction-level records, opinion polls or binary signals. Data can also be textual – extracted from books, documents, electronic media or text broadcasts in social networks, internet searches or Wikipedia, for example – or it can take the form of video, audio or global positioning systems records. *Whatever we do, our actions are apparently becoming more traceable and may soon be a new terrain for exploration and a traded commodity*.

Velocity; Technical developments have an impact on the speed at which data and information are generated and processed. New tools for processing and collecting, storing, mapping and linking micro-level data and new tools for analysing and extracting knowledge and patterns are constantly being developed and used in the

7 Mayer-Schoenberger, Viktor and Cukier, Kenneth, "Big Data: A Revolution That Will Transform How We Live, Work and Think", John Murray, 2013.

8 Kroes, Neelie, "Big Data for Europe", ICT 2013 Event – Session on Innovating by exploiting big and open data and digital content, Vilnius, 7 November 2013, http://europa.eu/rapid/press-release_SPEECH-13-893_en.htm

9 Hellerstein, Joe, University of California, Berkeley, quoted in, "Data, data everywhere", The Economist, 25 February 2010.

decision-making processes of business enterprises, governments and individuals. These new tools have an economic value as they are considered to give their users a competitive advantage. The ability to process large volumes of high frequency or tick data enables decision-makers to consider a significantly larger and more nuanced range of information in a shorter period of time. This additional information can help to identify the various probabilities associated with potential actions and reactions as part of taking decisions. This would increase the likelihood of taking decisions that are sounder, timelier and that can be reviewed more frequently.

Veracity; Accuracy and uncertainty surrounding data should be in the driving seat. The statistical principle of knowing your dataset before starting any analysis should prevail in these cases as well. Big data should be subject to the same quality standards and frameworks that apply to any other sources used for statistics. This is particularly the case regarding transparency in coverage and in the methodology used across regions and countries.

The speed at which relevant insights are extracted from big data, processed and managed seems to develop apace with the increasing availability and release of private and public micro-level sources of information. To keep abreast of the volume, variety, velocity and veracity of micro-level data, technical developments need to converge early enough and become standardised to facilitate the process of extracting and combining meaningful information, for instance from unstructured and non-hierarchical data or information stored in multiple sources. Nonetheless, history has shown that standardisation occurs only at a much later stage in the value chain, ie once the processing methods of combining sources and extracting information from multiple sources becomes a plain vanilla commodity with limited economic value.

A data service evolution? What seems clear is that the "industrial revolution of data" or the "social science data revolution"[10] is likely to constitute a "*data service evolution*". As it is becoming increasingly difficult for individuals to cope with huge volumes of constantly changing information, the risk is an explosion of often irrelevant, unclear or inaccurate information. This would leave consumers, professional users, public authorities and the general public struggling to find the means and time to distinguish between the quality and consistency of various sources. An increasing number of service providers will therefore be offering insights, intelligence and summary indicators using big data sources to address the need to transform detailed micro-data into summary information that is easier to understand, assess and communicate. Consequently, several corporations have set up "big data" strategies, universities are offering graduate programmes specifically tailored to the big data industry[11] and large amounts of venture capital are flowing into new specialised "big data" businesses.[12] In 2013 alone, approximately USD 4 billion was

10 King, Gary, "The Social Science Data Revolution", Department of Government at Harvard University, 30 March 2011: http://gking.harvard.edu/files/gking/files/evbase-horizonsp.pdf

11 Columbia University has launched online graduate programmes in this field: http://gsas.columbia.edu/news/columbia-university-launches-online-graduate-programs-meet-global-demand-big-data-careers, Harvard Business School has an executive programme: http://www.exed.hbs.edu/programs/data/Pages/default.aspx

12 Accel Partners has established funds specifically for big data driven software: www.accel.com/#companies

invested by US venture capital firms in supporting big data start-ups.[13] 2014 was the year of "big data" and 2015 was the year of the "Internet of Things", providing smart devices servicing consumers with information. The number of steps we take, the calories we burn, keeping track of our individual sleeping patterns, our purchases and consumption – all of these leave a digital trail. New devices connected to the internet increase our efficiency as well as our quality of life. According to one Gartner study[14], a typical family home could contain more than 500 smart devices by 2022, including sensing and remote control devices for media and entertainment, cookers, washing machines, heating and environmental measurements, security controls and voice recognition devices, representing an estimated $14 trillion market by 2022. Cisco Investments and Qihoo 360 Technology[15] have started investment funds of $150 million and $60 million respectively, targeting "Internet of Things" start-ups. Cisco indicates that in 2014 alone, venture capital firms invested USD 1.6 billion in such start-ups[16] – and this number is expected to grow in the coming years.

This "*data service evolution*" is leading to social and behavioural changes, with new cohesive societies being rebuilt and strengthened along thematic topics where interest groups are interacting via the internet and social media. These interactions are traceable, and can thus be monitored and converted into new services aimed at the respective communities (eg specific offers, loyalty programmes and collaborative opportunities). This will in turn affect the spending and saving behaviour and decision processes of households and firms. It is worth mentioning that, generally speaking, these cohesive societies have a positive effect on people's well-being, reflecting for instance the benefits of socialising online[17]. The benefits of social connections extend to people's probability of finding a job as well as their willingness to contribute voluntarily, share information and enhance the knowledge base of their network eg Wikipedia. People will switch social networks quickly and effortlessly as their values and interests change during the course of their lives.

However, as is detailed below, "big data" may also have drawbacks and become a source of concern for authorities such as central banks.

13 Datafloq sees itself as the "one-stop shop" for big data: https://datafloq.com/

14 http://www.gartner.com/newsroom/id/2839717

15 A Chinese Internet security company.

16 http://articles.economictimes.indiatimes.com/2014-10-15/news/55059521_1_iot-devices-internet

17 See OECD, "Perspectives on Global Developments 2012, Social cohesion in a shifting world", OECD, 2011, p. 202: http://dx.doi.org/10.1787/persp_glob_dev-2012-en, and Stiglitz, J. E, Sen, A. and Fitoussi, J., "Report by the Commission on the Measurement of Economic Performance and Social Progress", 2009.

2. Central banks and big data

While the availability and accessibility of large data sources is a new and rich field for statisticians, economists, econometricians and forecasters, it has been relatively unexploited for central banking purposes.[18]

As the mandates of many central banks have been extended to cover, in particular, financial stability and banking supervision in addition to monetary policy, this may suggest that the scope for using "big data" as a source of relevant information has increased. It would be of particular interest if such sources could help to detect trends and turning points within the economy, thereby providing supplementary and more timely information compared to the "traditional" toolkit of central banks.

Central banks already have access to a significantly large amount of statistics, intelligence, structured data and information, which are regularly fed into their decision-making process in the course of fulfilling their mandate. Indeed, central bankers are often viewed as "number crunchers", since they tend to process a tremendous amount of information in order to provide regular snapshots of the economy as well as to access and forecast economic developments over the short to medium-term horizon. In order to carry out those tasks, many techniques are applied to extract information from various parts of the economy. In addition, several models and methods have been developed by economists and researchers at central banks to facilitate this information extraction – eg static models, dynamic stochastic equilibrium models, ad hoc econometric estimation techniques – from large volumes of statistics.

Central banks therefore appear well-positioned to apply their existing models and econometric techniques to new datasets and/or to develop innovative methods to obtain timelier or new statistics and indicators. These supplementary statistics may provide further insights contributing to better guiding central bankers' policy actions as well as to assessing the subsequent impact and associated risks of these policy decisions on the financial system and real economy.

Big data could assist central bankers in obtaining a nearly real-time snapshot of the economy as well as providing early warning indicators to help identify turning points in the economic cycle. This may be particularly welcome in the light of the shortcomings observed in the run-up to the financial crisis.

In addition, there are new methods and techniques currently being developed by academic and private researchers to deal with new big data sources. For instance,

18 Vicente, María Rosalía, López-Menéndez, Ana J. and Pérez, Rigoberto, "Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing?", Technological Forecasting and Social Change, Volume 92, March 2015, pp. 132-139: http://www.sciencedirect.com/science/article/pii/S0040162514003904. D'Amuri, Francesco and Marcucci, Juri, "The predictive power of Google searches in forecasting unemployment", Economic Working Papers, Banca d'Italia, September 2012. Chadwick, Meltem Gülenay and Sengül, Gönül, "Nowcasting Unemployment Rate in Turkey: Let's Ask Google", Working Paper No 12/18, Central Bank of the Republic of Turkey, 2012.Artola, Concha and Galán, Enrique, "Tracking the future on the web: construction of leading indicators using internet searches", Documentos Ocasionales No 1203, Banco de España, September 2012.

text mining techniques open up new possibilities to assess what Keynes referred to as "animal spirits"[19] that cannot be captured in standard economic equations and quantitative variables. Sentiment indices harvested from internet articles, social media and internet search engines may, by applying adequate statistical algorithms, provide useful and timely insight into consumer sentiment, market uncertainty or systemic risk assessments.[20, 21, 22] Furthermore, new machine learning techniques and tools, such as support vector machines, non-linear random forest methods and elastic nets, are used to provide predictions on the basis of large, complex datasets.[23, 24]

Typically, statisticians, econometricians and researchers alike use well-established techniques (mainly relying on linear analysis) for various purposes, such as: (i) generating summary statistics and indicators; (ii) nowcasting/forecasting economic indicators; (iii) estimating missing data; and (iv) conducting hypothesis testing. However, machine learning techniques may be mobilised to find new and interesting patterns in large datasets, visualise such datasets, provide summary statistics and predictions and even generate new hypotheses and theories derived from the new patterns observed. This is because these new tools can help to identify non-linear relationships in datasets.

Another area where statistical tools are used is to assess and ensure the quality and fit of models on the basis of "in-sample" information; however, models may still perform less well in "out of sample" tests. Machine learning specialists are reported to have devised several ways to deal with this over-fitting problem faced by "traditional" models.

Moreover, social media are now well-established as a relatively new communication tool, bringing signals (eg financial tweets[25]) and various sentiment indicators to the attention of monetary authorities. In turn, social media are also gaining prominence in the central banking communication toolkit for sending clear and unfiltered messages to the public. This communication channel is amplified by followers re-tweeting messages, reaching an ever larger audience in a timely fashion.

19 Keynes, John Maynard, "The General Theory of Employment, Interest and Money", Palgrave Macmillan, 1936.

20 Daas, Piet J. H. and Puts, Marco J. H., "Social media sentiment and consumer confidence", ECB Statistics Paper Series, No 5, 2014.

21 Tobback, Ellen et al., "Belgian Economic Policy Uncertainty Index: Improvement through text mining", European Commission, paper presented at the ECB workshop on using big data for forecasting and statistics, Frankfurt am Main, April 2014.

22 Nyman, Rickard et al., "News and narratives in financial systems: exploiting Big Data for systemic risk assessment", Bank of England, paper presented at the ECB workshop on using big data for forecasting and statistics, Frankfurt am Main, April 2014.

23 Varian, Hal, "Big Data: New Tricks for Econometrics", 2014: http://people.ischool.berkeley.edu/~hal/Papers/2013/ml.pdf

24 Hastie, Trevor, Tibshirani, Robert and Friedman, Jerome, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Springer, 2nd edition, 2009: http://www-stat.stanford.edu/~tibs/ElemStatLearn/download.html

25 Cerchiello, Paola and Giudici, Paolo, "How to measure the quality of financial tweets", University of Pavia, paper presented at the ECB workshop on using big data for forecasting and statistics, Frankfurt am Main, April 2014.

Turning to banking supervision and regulation, there is a clear drive for regulatory authorities to obtain more micro-level data. [26, 27] Following the financial crisis, regulators have been keen to expand their data collection so as to better monitor financial risks and vulnerabilities. New big data sources may support these supervisory tasks; such sources include, for example, online operations in trading platforms, credit card payment transactions, mobile banking data, records related to securities settlement and cash payment systems, clearing houses and repurchase operations and derivatives settlement, as well as commercial and retail transactions and consumer internet purchases, to mention but a few.

This will give rise to new opportunities for regulators, particularly in Europe with the launch of the Single Supervisory Mechanism, but also more globally with the Financial Stability Board-led data collections on global systemically important banks (G-SIBs) and insurers (G-SIIs), with increasing demand for analysing micro-level data on counterparties and instruments.

The statistician George Box[28] is quoted as saying more than 30 years ago: "All models are wrong, but some are useful". This suggests that some combination of quantitative equations and theories of economic behaviour may help to capture and explain – to a certain extent – the world surrounding us.

Indeed, central bankers are trained to build models of how the world functions and test various hypotheses, and the results either confirm or reject theories based on assumptions of links between two or more variables. One important caveat is to recognise that correlation is not causation: no firm conclusion can be drawn simply on the basis of a correlation between two variables, as their correlations could be coincidental.

Compared with these traditional approaches, big data is bringing new ways of generating hypotheses, models and tests. With the availability of multiple large datasets, powerful computer processing and statistical algorithms, new patterns and causations may be found to generate new theories, in a sense reversing the traditional approaches when theories were set up and then confronted with the data (the top-down approach).

This new way of operating may challenge current theories. Big data may lead to a new "bottom-up" approach, whereby statistical algorithms and large covariance matrices using multiple sources generate new correlations and causations, resulting in new economic theories. The ability to apply cognitive and non-cognitive processing to quantitative information may better explain still relatively unknown behaviours, such as interconnectivity, herding effects, corrective actions and decision-making by financial agents.

Nonetheless, *it is human instinct to reject what cannot be (fully) comprehended. Instinct tends to signal danger whenever information growth outpaces our*

26 Nymand-Andersen, Per et al., "Financial data and risk information needed for the European System of Financial Supervision", Handbook of Financial Data and Risk Information I, Cambridge University Press 2014.

27 Bholat, David M., "The future of central bank data", Journal of Banking Regulation, No 14, June/August 2013.

28 George Edward Pelham Box was a British mathematician and professor of statistics at the University of Wisconsin. He was a pioneer of quality control, time series analysis, design of experiments and Bayesian inference.

understanding of how to process and digest it, and mind-sets will need to be changed, which will take time.

3. Data mania versus phobia

There is little doubt that new big data sources can be a rich playground for central banks. Big data will give central banks an opportunity to test existing theories, generate new correlations and causations among unexplored variables and pinpoint new signals among large amounts of noise. This is a paradise for both "hedgehogs" and "foxes", as it offers a means of making better predictions and stimulating academic and political debate. It creates an opportunity to adjust model-based theory and recognise the fragility of assumption-based models. The use of big data may make it easier to access knowledge from the feedback loops between economic and monetary forecasts, monetary policy implementation and market effects. Big data may also provide new opportunities to understand the behaviour involved in decision-making and to map network links among specific (financial) groups. They present an opportunity to extract additional detailed signals, contributing to more nuanced and varied views and thus fostering sound policy-making. From this perspective, big data may very likely have a welfare-enhancing effect on society.

All in all, big data may provide policy-makers with a broader and more in-depth range of economic and statistical indicators, and can do this more rapidly than traditional sources. Central banks, among others, may enhance forecasts by drawing on new data sources, using new techniques and integrating qualitative and quantitative information into their assessments, providing a more nuanced view for policy decisions.

Clearly, the quality of big data is an important issue: they should be subject to similar statistical quality standards as those that already prevail, such as transparency of sources, methodology, reliability and consistency over time. Big data are only of substantial value if they are both properly understood and appropriately analysed. Nonetheless, the new data sources may also lead to new approaches and models with regard to quality assessments.

Yet there remain caveats. One misperception of big data is the frequently expressed view that economists, econometricians and researchers do not need to worry about sample bias and representativeness, as large volumes of information will supersede standard sampling theory[29] given that the big data sources provide de facto census-type information.

For instance, access to all tweets would mean access to the characteristics of the entire "tweeting" population (corporates and members of the general public who have a Twitter account and send tweets). But the characteristics of this population may well differ from those people who do not tweet and are therefore excluded from the sample dataset. According to the Pew Research Center's "Social Networking Fact Sheet", 23% of online adults use Twitter, a figure that varies according to age, country,

29 Cukier, Kenneth and Mayer-Schoenberger, Viktor, "Big Data: A Revolution That Will Transform How We Live, Work and Think", Eamon Dolan, 2013.

education and ethnic origin.[30] Thus not all groups are represented or participate in this social medium. Additional information would therefore be needed to enable adjustments to be made and to gross figures up to the entire population if the aim is to extract signals and indicators on household sentiment or to start producing household indices using Twitter, for example. Moreover, statistical corrections will still have to be made for other features relating to unit measurements, double counting (tweeting and re-tweeting the same message), over-representativeness and over-fitting of models. In an event driven context – such as tweets or internet searches – volume changes does not necessary refer neither to reporting units nor to changes in demand. Take for instance the increased focus on "VW" – with the expected increases of internet searches and tweets related to VW. These potential increases may well be driven by the emission scandal of VW and not by an increased interest in purchasing cars. Therefore the data has to be adjusted if it is to be used as a leading indicator of car sales.

Other statistical challenges relate to the robustness and constancy of results over time, as many social media are event-based. Against this background, it is indeed misleading to conclude that sampling theory is not needed when working with large-volume big data sources. Attention has also been drawn to this misperception by the media[31]. On the contrary, statistical inferences are better understood when backed up by theory or at least by deeper thinking as to their potential causes. The Australian Bureau of Statistics is one of the pioneers studying interference in big data sources using a Bayesian framework.[32]

This is where statisticians will have an important role to play for central banks. Central banks take decisions intended to affect the behaviour and expectations of economic agents as part of the monetary transmission mechanism. One essential building block of central bankers' thinking is the "Lucas critique"[33], relating to that fact that the effects of macroeconomic policy decisions cannot be predicted entirely on the basis of historical correlations because the optimal decision rules of agents vary systematically in response to policy actions, typically by altering expectations. In other words, economic agent response functions change on the basis of policy actions and new information – this is also referred to as the feedback loop of policy actions.

Another area for statisticians' involvement is in balancing, through their experience, the need to ensure sufficient quality and users' desire to use big data in their search for rapid answers. Statistical quality cannot be taken for granted and needs to be taken seriously if statistics are to provide an accurate reflection of the structure and dynamics of our economies. Moreover, large datasets do not speak for themselves: they have to be described and contextualised before they can provide

30 Pew Research Center's Internet Project research, September 2014. "As of January 2014, 74% of online adults use social networking sites."

31 Harford, Tim, "Big data: are we making a big mistake?", Financial Times, 28 March 2014.

32 Tam, S.M. and Clarke, R., "Small steps towards Big Data – Some initiatives by the Australian Bureau of Statistics", International Statistical Review, presentation at the ECB workshop on using big data for forecasting and statistics, Frankfurt am Main, April 2014.

33 Lucas, R., "Econometric Policy Evaluation: A Critique", in Brunner, K. and Meltzer, A. "The Phillips Curve and Labor Markets", Carnegie-Rochester Conference Series on Public Policy, American Elsevier, New York, 1976, pp. 19-46.

useful insights. Similarly, it is important that the new big data sources are transparent in terms of their methodology and how the data are generated. Otherwise, the value of policy advice and forecasting using big data will be seriously undermined.

This raises the question of the policy measures needed to facilitate the work of statisticians, economists and researchers in their ability to pilot and apply these new sources as part of the policy tool-kit.

If we believe in the *inverse* of Goodhart's law,[34] how do we ensure that big data sources will remain available, accessible and free of charge as a public commodity if statistics offices, central banks and other public authorities start using these regularly as reliable sources? When the popularity of a big data source increases, it may become a pricey commercial asset, out of reach of authorities' budgets. This naturally raises questions of data ownership and the ethics of using big data. *Should big data sources on individual behaviours and patterns be commercialised or should they become a public commodity that complies with confidentiality and privacy rules – features that are jealously protected in other types of statistical areas*?

The availability of new data sources in public domains should be fostered. *The wealth of information and derived knowledge should be a public commodity* and made freely available at least to universities, researchers and government agencies.

4. Stay on the curve

It is important to realise that big data are not a particularly new phenomenon. For instance, helicopters are using infrared cameras to forecast demands for oil reserves, providing financial clients with reliable estimates for trading commodities in advance of official "oil demand" releases. Satellite images of car parks at large retailers and supermarket chains have been used to provide sales forecasts for trading purposes, in advance of public disclosure and quarterly reports, while satellite images are used for estimating the ability of order companies to produce and deliver large goods on time.

Big data are also used in the construction of official statistics. For example, scanner data are used in the production of consumer price indices (CPIs) by Statistics Norway to compute a sub-index for food and non-alcoholic beverages.[35] Similarly,

34 A theory introduced by Professor Charles Goodhart stating that when a measure becomes the target, it can no longer be used as the measure. Goodhart's law was originally applied to the stability of economic spending and is now used to point out the problem of assigning value to a specific variable to be used as an indicator.

Goodhart, C. A. E., "Problems of Monetary Management: The U.K. Experience", Papers in Monetary Economics, Reserve Bank of Australia, 1975.

Chrystal, K. A. and Mizen, P. D., "Goodhart's Law: Its Origins, Meaning and Implications for Monetary Policy", prepared for the Festschrift in honour of Charles Goodhart held on 15-16 November 2001 at the Bank of England.

35 Rodriguez, J. and Haraldsen, F., "The use of scanner data in the Norwegian CPI: The 'new' index for food and non-alcoholic beverages", Economic Survey 4, 2006, pp. 21-28.

Statistics Netherlands uses supermarket scanner data, [36] while the Swiss Federal Statistical Office has used prices taken from scanner data instead of the prices formerly collected in retail outlets when calculating its price indices.[37] It is easy to imagine satellite images being used to calculate large areas of farmland for the purposes of estimating more precisely crop production, the potential impact of natural disasters, as well as commodities demand and supply functions.

The Eurosystem is currently experimenting with these new data opportunities and has started regular collection of Google search data to nowcast certain macroeconomic variables on the basis of internet search terms. Other valuable sources may well include the "Billion Prices Project",[38] for example – particularly in countries where statistics are scarce or of lower quality.

The list of examples is seemingly endless. For instance, postal services data and tracing service devices can be used to estimate trade statistics and import and export tables across countries[39] and to try to estimate the impact of exchange rates.

Big data are a reality and while central bankers do not have to be ahead of the curve,[40] central banks will eventually need to take action and adopt a policy on how to deal with these new data sources.

Private service providers will continue to expand their releases of indicators based on big data sources, some of which may trace existing economic and financial statistics and indicators at a higher frequency and in advance of official releases. As users struggle to select relevant statistics and to find sufficient time to distinguish and evaluate between "good" and "not so good" statistics, they will tend to favour statistical sources that are readily understandable, accessible and easy to re-use. This may well present central banks and other suppliers of official statistics with a new communication challenge[41]. This relates to more than statistics and extends through the analysis, assessments and policy-decision chain.

A case in point is, for example, that of a supplier of big data on prices, such as Google, Amazon or eBay, with access to intraday or daily offering prices for consumer products via their platforms. It may be reasonable to expect such suppliers to be in a position to publish and commercialise a daily release of consumer prices or a sub-set. Although the methodology and quality of such an indicator may not meet the quality and transparency standards of official statistics, it may nevertheless be used and quoted by the press and re-used by the professional user community simply because

36 Van der Grient, H. and de Haan, J., "The use of supermarket scanner data in the Dutch CPI", Statistics Netherlands, 2010.

37 Becker-Vermeulen, C., "Recent developments in the Swiss CPI: scanner data, telecommunications and health price collection", paper presented at the ninth meeting of the Ottawa Group, London, 14-16 May 2006.

38 Cavallo, A., "The Billion Prices Project: Research and Inflation Measurement Applications", MIT Sloan School of Management, presentation at the ECB workshop on using big data for forecasting and statistics, Frankfurt am Main, April 2014.

39 Anson, J., Boffa, M. and Helble, M., "A short-run analysis of exchange rates and international trade", Universal Postal Union (UPU), paper presented at the ECB workshop on using big data for forecasting and statistics, Frankfurt am Main, April 2014.

40 According to Infochimps, 55% of big data projects are not completed: http://bigdata.infochimps.com

41 Nymand-Andersen, P., (2013) Safeguarding trust in statistics and the new statistical voice, Statistical journal of the International Association for Official Statistics 29, IOS Press, 2013

its frequency is high and it is easily accessible. If the data are sufficiently consistent over time, they may well be used as an early indicator to gauge the directions that will be published in official statistics, and will therefore feed into the decision-making process of economic agents and the general public before official policy decisions are taken. Official statistical agents and central banks may soon find themselves operating in a defensive mode. Given the ongoing dynamism of the economy, this can be a positive development too. For instance, big data sources are the driving force behind the renaissance of statistical algorithms, already increasingly applied in near real-time in algorithm trading, internet search engines, translation services and online sports betting, for instance. It is important for central banks to take advantage of these new opportunities by exploring their value in contributing to our understanding of the complex and interrelated components of financial markets and the real economy.

Central banks also need to join forces in exploring and assessing the usefulness of selective big data sources[42] and in defining a framework that can be used systematically as part of exploring their relevance for the central banking tool-kit. This would require a feasibility study and show-casing a few relevant sources. A road map of this kind would need to answer bold questions, such as:

(i) What relevant big data sources could be potentially valuable for central banking purposes and what types of insights could be extracted from these data sources?

(ii) Should central banks be actively involved in statistics algorithms, machine learning techniques, text mining and semantic analysis as part of extracting supplementary signals and intelligence from various aspects of the economy? And/or should central banks contribute by applying conventional models using big data sources?

(iii) What are the challenges for systematic, reliable and sustainable production of indicators based on these new data sources? Do we need new technology to cope with high frequency and large volumes variety and veracity of data, or should the central banking community develop partnerships with the source owners and associations? What methodology and quality challenges are inherent within relevant big data sources and how can they be measured as part of fitting or extending existing quality standards?

(v) Do central banks need to embark on data models and standardisation of micro-level data of big data sources and build bridges to international standards and classifications, thereby enabling supplementary sources/indicators to be integrated into existing statistics? What kind of governance structure is needed to ensure the sustainability of supply using big data sources?

(vi) What kind of skills, knowledge, resources and cost considerations are required to engage and drive down the big data avenue? Do central banks need to recruit candidates with data science skills or re-schooling existing staff?

42 See the Irving Fisher Committee on Central Bank Statistics report on "Central Banks use of and interest in big data". October 2015.

(vii) How are other public authorities approaching these challenges and what are the synergies and lessons learned from other pilots?

(viii) Should central banks take part in the growing public debate on the "ethics" of big data, protection of individual data, the methodology for aggregating micro-level data for ensuring representativeness in summary statistics and the regulation of big data sources? Should central banks engage in validating big data sources by providing methodological recommendations and transparently releasing these indicators and associated methodologies if they are relevant to the central banking tool-kit?

(x) What communication challenges and strategies are required for using big data sources and should central banks describe methodological differences, explaining these differences in private statistics released? How can quality and trust in these statistics and institutions be communicated and preserved?

The *big data service evolution* will continue to make progress. How quickly it will enter the realm of central banking policy will depend on how the wealth of existing statistical and economic knowledge within the central banking community can be proactively mobilised as part of entering the big data learning curve. The way forward may be to take small steps in developing and applying a structural approach for piloting the progressive use of big data, or non-official sources, for central banking purposes.

5. Conclusion

Big data are becoming an increasingly important aspect of our daily lives, both in connection with government policies and as new business models for corporates. They have been identified as providing one of the new services with a high growth potential. They use the digital footprint reflecting the way we live our lives, interact, communicate and socialise, where we shop, how we manage our health and wealth and the way we structure and organise our societies. The opportunities seem boundless as more and more private and government information becomes digital and publicly available.

The availability of big data, computer processing power, statistical algorithms and new patterns and causations from multiple large datasets may result in new theories being generated that may not have been found using traditional approaches. This new way of thinking challenges current theories. Big data may well lead to a new paradigm shift as a "bottom-up" approach, where statistical algorithms and large covariance matrices using multiple sources will generate new causations leading to new economic theories. As we gain the ability to merge qualitative and quantitative information, we may be able to better explain interconnectivity, herding effects, collective actions and behaviour by multiple actors irrespective of cultural differences and varying economic and financial market structures.

Big data offer central banks the opportunity to test existing theories, to generate new causations among unexplored variables and to detect new signals among large amounts of noise. They provide an opportunity to adjust model-based theory and to

recognise the fragility of models that are based on assumptions. They may facilitate the use of knowledge obtained from the feedback loops between monetary policy implementation and market effects. Exploring these new data sources could assist central bankers in obtaining a near real-time snapshot and potential early warning indicators. With regard to banking supervision and regulation, there is a clear drive to obtain more micro-level data. Regulators need to expand their data collection and to monitor systemic financial and banking risks and vulnerabilities.

The ability to extract timely summary statistics makes it possible for decision-makers at central banks to consider a significantly larger and more nuanced range of information in a shorter period of time and to obtain associated risk probabilities.

The new *big data service evolution* will continue to impact and transform our societies. However, big data are potentially of huge value if they are appropriately understood; otherwise, any results will merely contribute to the large amount of noise that already exists. The quality of the new data sources cannot be taken for granted and should be subject to similar transparency, quality standards and frameworks as those applied to official statistics.

While central banks clearly do not need to be ahead of the curve, this paper argues that central banks need to join forces and define a framework that can be used systematically as part of exploring its relevance for the tool-kit of central banks. This would require a feasibility study including piloting relevant big data sources and specifying a road map to position the central banking community on the big data curve.

This roadmap will lend support to the challenging decision making process and ensure that the central banking community stays tuned as we move into the next phases of the *big data service evolution*.

Bibliography

Anson, J., Boffa, M. and Helble, M., "A short-run analysis of exchange rates and international trade", Universal Postal Union (UPU), paper presented at the ECB workshop on using big data for forecasting and statistics, Frankfurt am Main, April 2014.

Artola, C. and Galán, E., "Tracking the future on the web: Construction of leading indicators using internet searches", *Occasional Paper No 1203*, Banco de España, September 2012.

Becker-Vermeulen, C., "Recent developments in the Swiss CPI: Scanner data, telecommunications and health price collection", paper presented at the ninth meeting of the Ottawa Group, London, 14-16 May 2006.

Bholat, D. M., "The future of central bank data", *Journal of Banking Regulation*, No 14, June/August 2013.

Box, G. E. P. and Draper N. R., "Empirical Model-Building and Response Surfaces", Wiley, 1987, p. 424.

Cavallo, A, "The Billion Prices Project: research and inflation measurement applications", MIT Sloan School of Management, paper presented at the ECB workshop on using big data for forecasting and statistics, Frankfurt am Main, April 2014.

Cerchiello, P. and Giudici P., "How to measure the quality of financial tweets", Quality and Quantity Volume 50, Issue 4, pp 1695–1713.

Chadwick, M. G. and Sengül. G., "Nowcasting unemployment rate in Turkey: Let's ask Google", *Working Paper No 12/18*, Central Bank of the Republic of Turkey, 2012.

Chrystal, K. A. and Mizen P. D., "Goodhart's Law: Its Origins, Meaning and Implications for Monetary Policy", prepared for the Festschrift in honour of Charles Goodhart held on 15-16 November 2001 at the Bank of England.

Mayer-Schoenberger, V. and Cukier, K., Big Data: "A Revolution That Will Transform How We Live, Work and Think", John Murray, 2013.

Daas, Piet J. H. and Puts, Marco J. H., "Social media sentiment and consumer confidence", *ECB Statistics Paper Series*, No 5, 2014.

D'Amuri, F. and Marcucci J., "The predictive power of Google searches in forecasting unemployment", *Economic Working Papers*, Banca d'Italia, September 2012.

Goodhart, C. A. E., "Problems of Monetary Management: The U.K. Experience", *Papers in Monetary Economics*, Reserve Bank of Australia, 1975.

Harford T., "Big data: are we making a big mistake?", *Financial Times*, 28 March 2014.

Hastie, T., Tibshirani, R. and Friedman, J., "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Springer, 2nd edition, 2009: http://www-stat.stanford.edu/~tibs/ElemStatLearn/download.html

Infochimps, a CSC big data business website: https://www.infochimps.com/

Irving Fisher Committee on Central Bank Statistics, "Central Banks use of and interest in big data", October 2015. http://www.bis.org/ifc/publ/ifc-report-bigdata.pdf

Keynes, J. M., "The General Theory of Employment, Interest and Money", Palgrave Macmillan, 1936.

King, G., "The Social Science Data Revolution", Department of Government at Harvard University, 30 March 2011: http://gking.harvard.edu/files/gking/files/evbase-horizonsp.pdf

Kroes, N., "Big data for Europe", ICT 2013 Event – Session on Innovating by exploiting big and open data and digital content, Vilnius 7 November 2011: http://europa.eu/rapid/press-release_SPEECH-13-893_en.htm

Laney, D., "3D Data Management: Controlling Data Volume, Velocity, and Variety", META Group (now Gartner), 6 February 2001: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lucas, R., "Econometric Policy Evaluation: A Critique", in Brunner, K. and Meltzer, A., "The Phillips Curve and Labor Markets", *Carnegie-Rochester Conference Series on Public Policy*, American Elsevier, New York, 1976, pp. 19-46.

Nyman, R. et al., "News and narratives in financial systems: Exploiting Big Data for systemic risk assessment", Bank of England, paper presented at the ECB workshop on using big data for forecasting and statistics, Frankfurt am Main, April 2014.

Nymand-Andersen, P., "Safeguarding trust in statistics and the new statistical voice", Statistical journal of the International Association for Official Statistics 29, IOS Press, 2013.

Nymand-Andersen, P., Antoniou N., Burkart O., Kure J. "Financial data and risk information needed for the European System of Financial Supervision", Handbook of Financial Data and Risk Information I: Principles and Context, Edited by Brose, M.S, Flood, M.D., Krishna, D., Nichols, B. Cambridge University Press. 2014.

OECD, "Perspectives on Global Developments 2012, Social cohesion in a shifting world", OECD, 2011, p. 202: http://dx.doi.org/10.1787/persp_glob_dev-2012-en.

Rodriguez, J. and Haraldsen, F., "The use of scanner data in the Norwegian CPI: The 'new' index for food and non-alcoholic beverages", *Economic Survey 4*, 2006, pp. 21-28.

Stiglitz, J. E, Sen, A. and Fitoussi, J., "Report by the Commission on the Measurement of Economic Performance and Social Progress", 2009

Pew Internet Project research, "Social Networking Fact Sheet", September 2014.

Tam, S. M. and Clarke, R., "Small steps towards Big Data – Some initiatives by the Australian Bureau of Statistics", *International Statistical Review*, presentation at the ECB workshop on using big data for forecasting and statistics, Frankfurt am Main, April 2014.

The Economist, "Data, data everywhere", 25 February 2010.

Tobback, E. et al., "Belgian Economic Policy Uncertainty Index: Improvement through text mining", European Commission, paper presented at the ECB workshop on using big data for forecasting and statistics, Frankfurt am Main, April 2014.

Van der Grient, H. and de Haan, J., "The use of supermarket scanner data in the Dutch CPI", Statistics Netherlands, 2010.

Varian, H., Big Data: "New Tricks for Econometrics", University of California, Berkeley, June 2013, revised in April 2014: http://people.ischool.berkeley.edu/~hal/Papers/2013/ml.pdf

Vicente, M. R., López-Menéndez, A. J. and Pérez, R., "Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing?", *Technological Forecasting and Social Change*, Volume 92, March 2015, pp. 132-139.

Annex I: Big data – references in the literature

Year	Reference	Definition
2015	Oxford Dictionaries[43]	"*Extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions*: *'much IT investment is going towards managing and maintaining big data'.*"
2014	Gartner[44]	"*Big data is **high-volume**, **high-velocity** and **high-variety** information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.*" This definition of big data is referred to as a **3V model**.
2014	IBM[45,46]	*"Every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big data."*
2014	SAS[47]	*"For most organizations, big data is the reality of doing business. It's the proliferation of structured and unstructured data that floods your organization on a daily basis – and if managed well, it can deliver powerful insights."*
2013	Fedtech Magazine[48]	*"New trends in IT are often thought of in terms of leading-edge technology solutions to significant enterprise challenges. In other words, organizations face challenges (for example, improving data center efficiency or providing IT services to remote workers) and IT solutions (such as cloud computing and mobile devices) address them. But one of the biggest trends in IT today, Big Data, is actually named for the challenge it represents, rather than the solution. At its core, Big Data means lots of data — so much data collected via so many evolving mechanisms that it can be overwhelming. It is increasingly easy for government agencies to store all that data. But Big Data as an IT strategy requires making sense of the data being collected — processing, analyzing and exploiting it for government, partner and constituent gain.* *Big Data refers to digital information that is massive and varied, and that arrives in such waves that it requires advanced technology and best practices to sort, process, store and analyze. Organizations that do so effectively can use it to their advantage. Big Data is less about the terabytes than it is about the query tools and business intelligence software needed to make sense of the terabytes."*
2012	Microsoft[49]	*"The increasingly large and complex data that is now challenging traditional database systems YouTube videos, Facebook posts, credit card transactions, store inventory, your last grocery purchase. Trillions of pieces of information are being collected, stored, and analyzed almost daily with increasing speed. Big Data addresses one of the most critical issues facing business today: how to gain value from the growing reams of complex data."*

43 Oxford Dictionaries: http://www.oxforddictionaries.com/definition/english/big-data

44 Gartner: http://www.gartner.com/it-glossary/big-data/

45 IBM: http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

46 IBM: Big data, bigger outcomes (PDF)

47 SAS: http://www.sas.com/en_us/insights/big-data.html

48 Fedtech: http://www.fedtechmagazine.com/sites/default/files/122210-wp-big-data-df.pdf

49 Microsoft: http://download.microsoft.com/download/ ... /Microsoft_Big_Data_Booklet.pdf

2012	Global Pulse[50]	"*'Big Data' is a popular phrase used to describe a massive volume of both structured and unstructured data that is so large that it's difficult to process with traditional database and software techniques. The characteristics which broadly distinguish Big Data are sometimes called the '3 V's': more volume, more variety and higher rates of velocity. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos posted online, transaction records of online purchases, and from cell phone GPS signals to name a few. This data is known as 'Big Data' because, as the term suggests, it is huge in both scope and power.*"
2011	McKinsey[51]	"*'Big data' refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze. This definition is intentionally subjective and incorporates a moving definition of how big a dataset needs to be in order to be considered big data—ie, we don't define big data in terms of being larger than a certain number of terabytes (thousands of gigabytes). We assume that, as technology advances over time, the size of datasets that qualify as big data will also increase. Also note that the definition can vary by sector, depending on what kinds of software tools are commonly available and what sizes of datasets are common in a particular industry. With those caveats, big data in many sectors today will range from a few dozen terabytes to multiple petabytes (thousands of terabytes).*"
2001	META Group[52] (now Gartner)	Gartner's paper does not use the term "big data" but rather reports on data management challenges: "*While enterprises struggle to consolidate systems and collapse redundant databases to enable greater operational, analytical, and collaborative consistencies, changing economic conditions have made this job more difficult. E-commerce, in particular, has exploded data management challenges along three dimensions:* ***volumes****,* ***velocity****, and* ***variety****.*" This "**3V model"** has almost become the standard way of defining big data.

50 Global Pulse: Big Data for Development: Challenges & Opportunities

51 McKinsey: Big data: The next frontier for innovation, competition, and productivity (see full report as pdf).

52 Laney, Douglas: "3D Data Management: Controlling Data Volume, Velocity, and Variety", Gartner, 2001.

Annex II: A few illustrative examples on the use of big data

Example 1	
Companies	Progressive, Tesco Bank, Generali
Industry	Insurance
Big data	Data comes from a "telematics device", which has been installed in a car. The device monitors driving data: speed, braking habits, etc. For example, Progressive has more than a trillion seconds of driving data from 1.6 million customers.
Predict	Companies offering car insurance try to assess the riskiness of a driver based on the data received from telematics devices.
Business model	New customers are acquired by recognising low risk drivers and offering them cheaper car insurance. According to Mike Fitzgerald, Senior Analyst at Celent, most people are overpaying.
References	BBC News: How big data is changing the cost of insurance SAS: Telematics: How Big Data Is Transforming the Auto Insurance Industry

Example 2	
Company	PASSUR Aerospace
Industry	Aviation business intelligence company
Big data	Weather, flight schedules, radar about aircraft in the local sky.
Predict	Calculate estimated time of arrival of an aeroplane using big data. An airline company can reduce costs if it has information on accurate landing times.
Business model	Web-based solutions for airlines.
References	HBR: Big Data's Management Revolution

Example 3	
Company	Target
Industry	Second-largest discount retailer in the United States after Walmart
Big data	Shopping cart history and any demographic information collected.
Predict	Find out whether a customer has a baby, and focus on marketing baby products.
Business model	Build customer loyalty by knowing what products the customer needs.
References	Forbes: How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did The New York Times: How Companies Learn Your Secrets

Example 4	
Company	Google
Industry	Internet
Big data	Google search data
Predict	Use Google queries to predict economic activity. Google released a research paper in which they show that the volume of Google searches is an economic indicator. In the paper, Google's Chief Economist, Hal Varian, gives an example of Ford vehicle sales. Apparently, when searches for the keyword "Ford" decrease, so do Ford vehicle sales.
Business model	Make Google search data more valuable.
References	Google: Predicting the Present with Google Trends

Example 5	
Company	UPS
Industry	Courier
Big data	Vehicles are equipped with telematics devices, which monitor over 200 elements: speed, RPM, oil pressure, seat belt use, number of times the truck is placed in reverse, idling time, etc.
Predict	By analysing the data, UPS can determine the condition of their trucks and also monitor and improve driving habits. Knowing the condition of a truck is important for reducing maintenance costs. Maintenance at regular intervals is not needed – maintenance is carried out when needed.
Business model	Cost reduction
References	Automotive fleet: Telematics Sensor-Equipped Trucks Help UPS Control Costs

Example 6	
Companies	Genscape, RS Metrics
Industry	Satellite imagery and geospatial information
Big data	Satellite images and pictures taken from a helicopter using a heat-sensitive camera.
Predict	Genscape has been taking pictures of oil tanks to determine how full they are. Lowering oil level in tanks might be a sign of future demand. RS Metrics has analysed images of parking lots to determine shopper traffic compared to previous years.
Business model	Sell information to those who have financial interest in the matter.
References	The Wall Street Journal: Traders Seek an Edge With High-Tech Snooping

Example 7	
Companies	Zions Bank
Industry	Banking
Big data	140 data sources, including core banking, online banking and loan servicing data.
Predict	For example, they can detect if a customer makes a branch transaction at the same time as a mobile banking transaction.
Business model	Fraud prevention
References	Bankinfosecurity.com: Using Big Data to Prevent Fraud American Banker: How Zions Bank Is Conquering Big Data for Marketing Campaigns

Example 8	
Companies	IBM & Columbia University Medical Center
Industry	Software & Medical
Big data	Physiological data streams such as EEG feeds, blood pressure, blood oxygen levels and temperature readings in conjunction with persistent data, such as lab test results, patient information and symptoms reported by medical professionals and patients.
Predict	Researchers have been able to detect severe complications in brain-injured patients up to 48 hours earlier than by using traditional methods.
Business model	The technology comes from IBM and the need comes from the Medical Center.
References	IBM: IBM Analytics Helps Medical Researchers Detect Complication In Stroke Patients

Example 9	
Company	Ushahidi
Industry	Software
Big data	Text messages, tweets from Twitter, photographs, video and web-based reports.
Predict	Ushahidi has a platform which can draw a crisis map of humanitarian needs. The US Marine Corps was able to save hundreds of lives at the time of Haiti's earthquake by using Ushahidi's crisis map.
Business model	Non-profit humanitarian technology solutions
References	Forbes: Crisis Maps: Harnessing the Power of Big Data to Deliver Humanitarian Assistance

Example 10		
Company	Georgia State Board of Pardons and Paroles	
Industry	Government	
Big data	Information on prisoners and parolees: age at first offence, gender, attitude and behaviour during supervision or employment, etc. Forty-five risk factors have been detected, helping to predict the likelihood of a parolee committing another crime.	
Predict	It costs around USD 51 a day to maintain a single prisoner, and USD 2.90 to supervise a parolee. However there is a risk: 30% of parolees will commit another crime or otherwise break the terms of their parole and return to prison. With big data the State Board can identify the prisoners in the 70% share.	
Business model	Cost reduction	
References	Federal Reserve Bank of Atlanta:	Big Data: Government's Next Frontier