CAPCO

THE CAPCO INSTITUTE JOURNAL OF FINANCIAL TRANSFORMATION

DATA ANALYTICS

Using big data analytics and artificial intelligence: A central banking perspective

OKIRIZA WIBISONO | HIDAYAH DHINI
ARI ANGGRAINI WIDJANARTI
ALVIN ANDHIKA ZULEN | BRUNO TISSOT

DATA ANALYTICS

50TH EDITION | NOVEMBER 2019

THE CAPCO INSTITUTE

JOURNAL OF FINANCIAL TRANSFORMATION

RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

CONTENTS

DATA MANAGEMENT

DATA ANALYTICS

DATA INTELLIGENCE

DEAR READER,

Welcome to the milestone 50th edition of the Capco Institute Journal of Financial Transformation.

Launched in 2001, the Journal has covered topics which have charted the evolution of the financial services sector and recorded the fundamental transformation of the industry. Its pages have been filled with invaluable insights covering everything from risk, wealth, and pricing, to digitization, design thinking, automation, and much more.

The Journal has also been privileged to include contributions from some of the world's foremost thinkers from academia and the industry, including 20 Nobel Laureates, and over 200 senior financial executives and regulators, and has been co-published with some of the most prestigious business schools from around the world.

I am proud to celebrate reaching 50 editions of the Journal, and today, the underlying principle of the Journal remains unchanged: to deliver thinking to advance the field of applied finance, looking forward to how we can meet the important challenges of the future.

Data is playing a crucial role in informing decision-making to drive financial institutions forward, and organizations are unlocking hidden value through harvesting, analyzing and managing their data. The papers in this edition demonstrate a growing emphasis on this field, examining such topics as machine learning and AI, regulatory compliance, program implementation, and strategy.

As ever, you can expect the highest caliber of research and practical guidance from our distinguished contributors, and I trust that this will prove useful to your own thinking and decision making. I look forward to sharing future editions of the Journal with you.

Lance Levy, **Capco CEO**

FOREWORD

Since the launch of the Journal of Financial Transformation nearly 20 years ago, we have witnessed a global financial crisis, the re-emergence of regulation as a dominant engine of change, a monumental increase in computer processing power, the emergence of the cloud and other disruptive technologies, and a significant shift in consumer habits and expectations.

Throughout, there has been one constant: the immense volume of data that financial services institutions accumulate through their interactions with their clients and risk management activities. Today, the scale, processing power and opportunities to gather, analyze and deploy that data has grown beyond all recognition.

That is why we are dedicating the 50th issue of the Journal of Financial Transformation to the topic of data, which has the power to change the financial industry just as profoundly over the coming 20 years and 50 issues. The articles gathered in this issue cover a broad spectrum of data-related topics, ranging from the opportunities presented by data analytics to enhance business performance to the challenges inherent in wrestling with legacy information architectures. In many cases, achieving the former is held back by shortcomings around the quality of, and access to, data arising from the latter.

It is these twin pillars of opportunity and challenge that inform the current inflection point at which the financial industry now stands. Whilst there is opportunity to improve user experiences through better customer segmentation or artificial intelligence, for example, there are also fundamental challenges around how organizations achieve this – and if they can, whether they should.

The expanding field of data ethics will consume a great deal of senior executive time as organizations find their feet as they slowly progress forward into this new territory. In my view, it is critical that organizations use this time wisely, and do not just focus on short-term opportunities but rather ground themselves in the practical challenges they face. Financial institutions must invest in the core building blocks of data architecture and management, so that as they innovate, they are not held back, but set up for long-term success.

I hope that you enjoy reading this edition of the Journal and that it helps you in your endeavours to tackle the challenges of today's data environment.

Guest Editor
Chris Probert, **Partner, Capco**

USING BIG DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE: A CENTRAL BANKING PERSPECTIVE

OKIRIZA WIBISONO | Big Data Analyst, Bank Indonesia
HIDAYAH DHINI ARI | Head of Digital Data Statistics and Big Data Analytics Development Division, Bank Indonesia
ANGGRAINI WIDJANARTI | Big Data Analyst, Bank Indonesia
ALVIN ANDHIKA ZULEN | Big Data Analyst, Bank Indonesia
BRUNO TISSOT | Head of Statistics and Research Support, BIS, and Head of the IFC Secretariat[1]

ABSTRACT

Information and the internet technology have fostered new web-based services that affect every facet of today's economic and financial activity. For their part, central banks face a surge in "financial big datasets", reflecting the combination of new, rapidly developing electronic footprints as well as large and growing financial, administrative, and commercial records. This phenomenon has the potential to strengthen analysis for decision making, by providing more complete, immediate, and granular information as a complement to "traditional" macroeconomic indicators. To this end, a number of techniques are being developed, often referred to as "big data analytics" and "artificial intelligence". However, getting the most out of these new developments is no trivial task. Central banks, like other public authorities, face numerous challenges, especially in handling these new data and using them for policy purposes. This paper covers three main topics discussing these issues: the main big data sources and associated analytical techniques that are relevant for central banks, the type of insights that can be provided by big data, and how big data is actually used in crafting policy.

1. INTRODUCTION

Information and the internet technology have fostered new web-based services that affect every facet of today's economic and financial activity. This creates enormous quantities of "big data" – defined as "the massive volume of data that is generated by the increasing use of digital tools and information systems." [FSB (2017)]. Such data are produced in real time, in differing formats, and by a wide range of institutions and individuals. For their part, central banks face a surge in "financial big datasets," reflecting the combination of new, rapidly developing electronic footprints as well as large and growing financial, administrative, and commercial records.

This phenomenon has the potential to strengthen analysis for decision making, by providing more complete, immediate, and granular information as a complement to "traditional" macroeconomic indicators. To this end, a number of techniques are being developed, often referred to as "big data analytics" and "artificial intelligence" (AI). These promise faster, more holistic, and more connected insights, as compared with

[1] The views expressed here are those of the authors and do not necessarily reflect those of Bank Indonesia, the Bank for International Settlements (BIS), or the Irving Fisher Committee on Central Bank Statistics (IFC). This article draws on the various presentations made on the occasion of the workshop on "Big data for central bank policies" and the high-level policy-oriented seminar on "Building pathways for policymaking with big data" organized by Bank Indonesia with support from the BIS and the IFC in July 2018 at Bali, Indonesia. The proceedings were published in the IFC Bulletin series [IFC (2019)].

traditional statistical techniques and analyses. An increasing number of central banks have launched specific big data initiatives to explore these issues. They are also sharing their expertise in collecting, working with, and using big data, especially in the context of the BIS's Irving Fisher Committee on Central Bank Statistics (IFC) [IFC (2017a)].

Getting the most out of these new developments is no trivial task for policymakers. Central banks, like other public authorities, face numerous challenges, especially in handling these new data and using them for policy purposes. In particular, significant resources are often required to handle large and complex datasets, while the benefits of such investments are not always clear-cut. For instance, to what extent should sophisticated techniques be used to deal with this type of information? What is the added value over more traditional approaches, and how should the results be interpreted? How can the associated insights be integrated into current decision-making processes and be communicated to the public? And, lastly, what are the best strategies for central banks seeking to realize the full potential of new big data information and analytical tools, considering, in particular, resource constraints and other priorities?

This paper covers three main areas that can shed light on these various questions. First, the main big data sources and associated analytical techniques that are relevant for central banks. Second, the types of insight that can be provided by big data from their perspective. And, third, a review of how big data is actually used in crafting central bank policies.

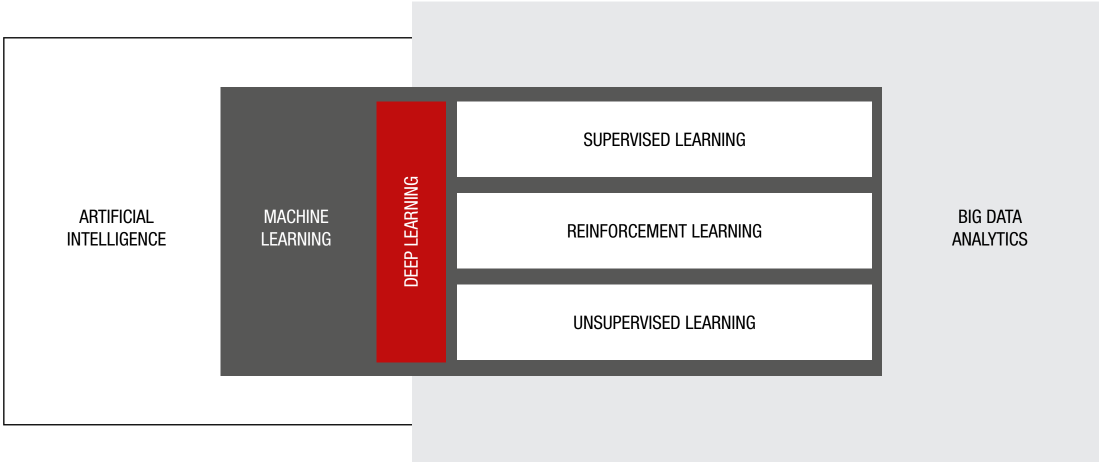
2. THE BIG DATA REVOLUTION: NEW DATA SOURCES AND ANALYTICAL TECHNIQUES

It is widely acknowledged that policymakers should not miss out on the opportunities provided by big data – described by some as the new oil of the 21st century [Economist (2017)]. Public institutions are not the main producers of big datasets, and some of this information may have little relevance for their daily work. Yet, central banks are increasingly dealing with "financial big data" sources that impinge on a wide range of their activities.

Data volumes have surged hand-in-hand with the development of specific techniques for their analysis, thanks to "big data analytics" – broadly referring to the general analysis of these datasets – and "artificial intelligence" (AI) – defined as "the theory and development of computer systems able to perform tasks that traditionally have required human intelligence" [FSB (2017)]. Strictly speaking, these two concepts can differ somewhat (for instance, one can develop tools to analyze big datasets that are not based on AI techniques), as shown in Figure 1.

In practice, big data analytics are not very different from traditional econometrical techniques, and indeed they borrow from many long-established methodologies and tools developed for general statistics; for instance, principal component analysis, developed at the beginning of the last century. Yet, one key characteristic is that they are applied to modern datasets that can be both very large and

Figure 1: A schematic view of AI, machine learning, and big data analytics



Source: FSB (2017)

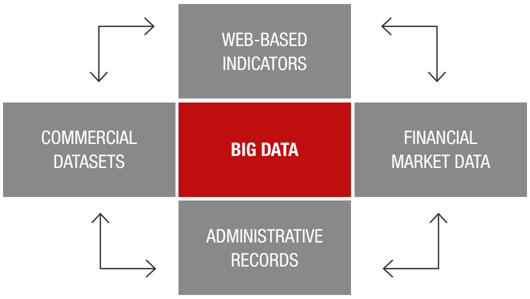
complex. Extracting relevant information from these sources is not straightforward, often requiring a distinct set of skills, depending on the type of information involved. As a result, big data analytics and AI techniques comprise a variety of statistical/modeling approaches, such as machine learning, text-mining techniques, network analysis, agent-based modeling,[2] etc.

The experience accumulated in recent years underlines that, indeed, there are specific big data sources of relevance to central banks. It also shows that a number of techniques developed for analyzing big data can play a useful role.

2.1 Big data information for central banks

Three main sources of big data are commonly identified.[3] These categories are related to (i) social networks (human-sourced information, such as blogs and searches); (ii) traditional business systems (process-mediated data, such as files produced by commercial transactions, e-commerce, credit card operations); and (iii) the internet of things (machine-generated data, such as information produced by pollution/traffic sensors, mobile phones, computer logs, etc). These are very generic categories, and, in practice, big data will comprise multiple types of heterogeneous datasets derived from these three main sources.

Figure 2: Four main types of financial big dataset



Focusing more specifically on central banks, four types of datasets would usually be described as financial big data (see Figure 2): internet-based indicators, commercial datasets, financial market indicators, and administrative records.[4]

Compared with the private sector,[5] central banks' use of web-based indicators may be somewhat more limited, especially with regards to unstructured data, such as images. Even so, several projects are under way to make use of data collected on the internet to support monetary and financial policymaking. Moreover, an important aspect relates to the increased access to digitalized information, reflecting both the fact that more and more textual information is becoming available on the web (e.g., social media) and also that "traditional" printed documents can now be easily digitalized, searched, and analyzed in much the same way as web-based indicators.

In reality, however, the bulk of the financial big datasets relevant to central banks consists of the very granular information provided by large and growing records covering commercial transactions, financial market developments, and administrative operations. This type of information has been spurred by the expansion of the micro-level datasets collected in the aftermath of the Great Financial Crisis (GFC) of 2007–09, especially in the context of the Data Gaps Initiative (DGI) endorsed by the G20 [FSB-IMF (2009)]. For instance, significant efforts have been made globally to compile large and granular loan-by-loan and security-by-security databases, as well as records of individual derivatives trades [IFC (2018)]. There has also been an increasing attempt by central banks to make a greater use of granular information covering firms' individual financial statements [IFC (2017b)]. As a result of these various initiatives, central banks now have at their disposal very detailed information on the financial system, including at the level of specific institutions, transactions, or instruments.

2.2 Extracting knowledge from large quantitative datasets: Classification and clustering

The expansion of big data sources has gone hand-in-hand with the development of new analytical tools to deal with them. The first, and particularly important, category of these big data techniques aims at extracting summary information from large quantitative datasets. This is an area that is relatively close to "traditional statistics", as it does not involve the treatment of unstructured information (e.g., text, images). In fact, many big datasets are well structured, and can be appropriately dealt with using statistical algorithms developed for numerical datasets. The main goal is to obtain summary indicators by condensing the large amount of data points available, basically by finding similarities between them (through classification) and regrouping them (through clustering).

[2] See Haldane (2018), who argues that big data can facilitate policymakers' understanding of economic agents' reactions through the exploration of behaviors in a "virtual economy".
[3] Following the work conducted under the aegis of the United Nations [see Meeting of the Expert Group on International Statistical Classifications (2015)].
[4] For the use of administrative data sources for official statistics, see for instance Bean (2016) in the U.K. context.
[5] Especially the major U.S. technology companies (GAFAs): Google, Apple, Facebook, and Amazon.

Many of these techniques involve so-called "machine learning". This is a subset of AI techniques, which can be defined as "a method of designing a sequence of actions to solve a problem that optimize automatically through experience and with limited or no human intervention" [FSB (2017)]. This approach is quite close to conventional econometrics, albeit with three distinct features. First, machine learning is typically focused on prediction rather than identifying a causal relationship. Second, the aim is to choose an algorithm that fits with the actual data observed, rather than with a theoretical model. Third, and linked to the previous point, the techniques are selected by looking at their goodness-to-fit, and less at the more traditional statistical tests used in econometrics.

There are several categories of machine learning, which can be split into two main groups. First, in "supervised machine learning", "an algorithm is fed a set of 'training' data that contain labels on the observations" [FSB (2017)]. The goal is to classify individual data points, by identifying, among several classes (i.e., categories of observations), the one to which a new observation belongs. This is inferred from the analysis of a sample of past observations, i.e., the training dataset, for which their group (category) is known. The objective of the algorithm is to predict the category of a new observation, depending on its characteristics. For instance, to predict the approval of a new loan ("yes" or "no", depending on its features as compared to an observed historical dataset of loans that have been approved or rejected); or whether a firm is likely to default in a few months. Various algorithms can be implemented for this purpose, including logistic regression techniques, linear discriminant analysis, Naïve Bayes classifier, support vector machines, k-nearest neighbors, decision trees, random forest, etc.

The second group is "unsupervised machine learning", for which "the data provided to the algorithm do not contain labels" [FSB (2017)]. This means that categories have not been identified ex-ante for a specific set of observations, so that the algorithm has to identify the clusters, regrouping observations for which it detects similar characteristics or "patterns". Two prominent examples are clustering and dimensionality reduction algorithms. In "clustering", the aim is to detect the underlying groups that exist in the granular dataset – for example, to identify groups of customers or firms that have similar characteristics – by putting the most similar observations in the same cluster in an agglomerative way (bottom-up approach).[6] "Dimensionality reduction" relates to the rearrangement of the original information in a smaller number of pockets, in a divisive (top-down) way; the objective is that the number of independent variables becomes (significantly) smaller, without too much compromise in terms of information loss.

There are, of course, additional types of algorithms. One is "reinforcement learning", which complements unsupervised learning with additional information feedback; for instance, through human intervention. Another is "deep learning" (or artificial neural networks), based on data representations inspired by the function of neurons in the brain. Recent evaluations suggest that deep learning can perform better than traditional classification algorithms when dealing with unstructured data, such as texts and images – one reason being that applying traditional quantitative algorithms is problematic, as it requires unstructured information to be converted into a numerical format. In contrast, deep learning techniques can be used to deal directly with the original raw data.

In view of this diversity, the choice of a specific algorithm will depend on the assumptions made regarding the features of the dataset of interest – for instance, a Naïve Bayes classifier would be appropriate when the variables are assumed to be independent and follow a Gaussian distribution. In practice, data scientists will have to identify which algorithm works best for the problem at hand, often requiring a rigorous and repetitive process of trial and error; this is often as much art as it is science.

In choosing the right "model", it is important to define an evaluation metric. The aim is to measure how well a specific algorithm fits, and to compare the performance of alternative algorithms. The most straightforward metric for classification is "prediction accuracy", which is simply the percentage of observations for which the algorithm predicts the class variable correctly (this will usually be done by comparing the result of the algorithm with what a human evaluator would conclude on a specific data sample). But an accuracy metric may not be suitable for all exercises, particularly in the case of an unbalanced distribution of classes. For example, when looking at whether a transaction is legitimate or fraudulent, a very simplistic model could be adopted that assumes that all transactions are legitimate: its accuracy will look very high,

[6] More precisely, cluster analysis can be defined as "a statistical technique whereby data or objects are classified into groups (clusters) that are similar to one another but different from data or objects in other clusters." [FSB (2017)].

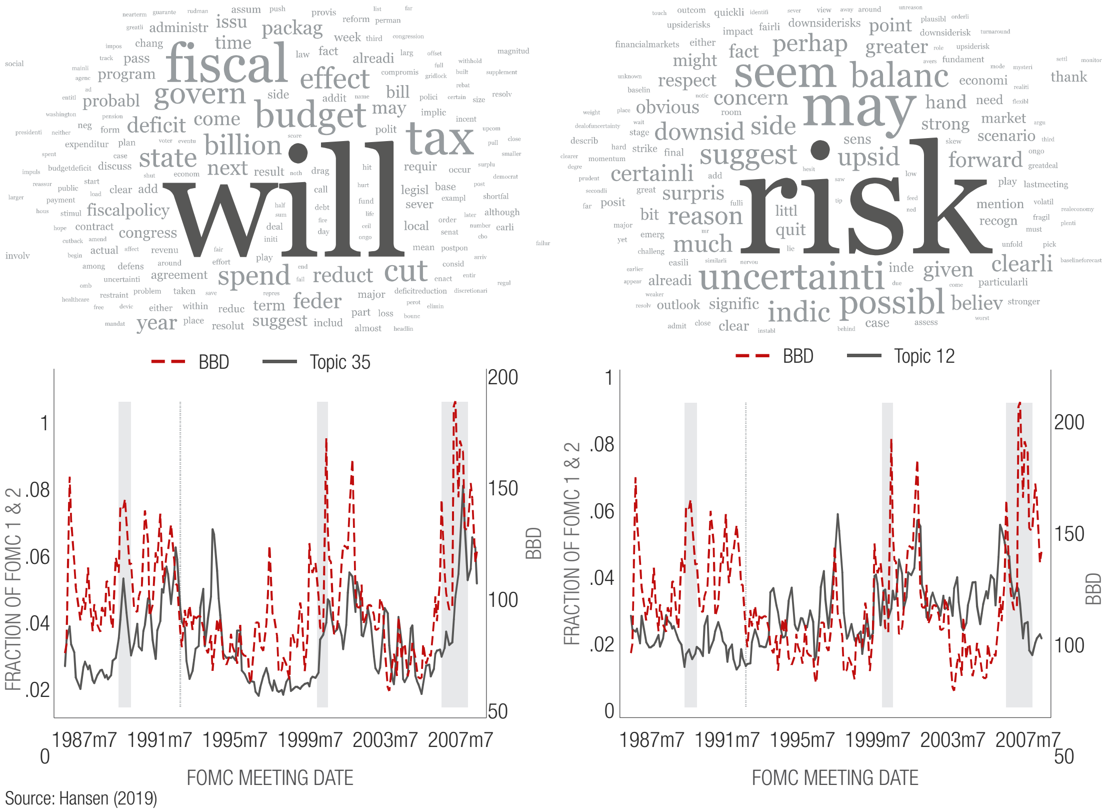
because a priori most transactions are not fraudulent; but the usefulness of such a simple model would be quite limited. Hence, other metrics have to be found for evaluating algorithms when the distribution of classes is highly unbalanced.[7] Another possible approach is to address the class imbalance issue at the observation level; for instance, by duplicating (over-sampling) elements from the minority class or, conversely, by discarding those (under-sampling) from the dominant class.

2.3 Text mining

Another rapidly developing area of big data analytics is text-mining, i.e., analysis of semantic information – through the automated analysis of large quantities of natural language text and the detection of lexical or linguistic patterns with the aim of extracting useful insights. While most empirical work in economics deals with numerical indicators, such as prices or sales data, a large and increasing amount of textual information is also generated by economic and financial activities – including internet-based activities (e.g., social media posts), but also the wider range of textual information provided by, say, company financial reports, media articles, public authorities' deliberations, etc. Analyzing this unstructured information has become of key interest to policymakers, not least in view of the important role played by "soft" indicators such as confidence and expectations during the GFC. And, indeed, text-mining techniques can usefully be applied to dealing with these data in a structured, quantitative way.

Text analysis typically starts with some standard "pre-processing steps", such as tokenization (splitting text into words), stopword removal (discarding very frequent/non-topical words, e.g., "a", "the", "to"), stemming or lemmatizing (converting words into their root forms, for instance, "prediction" and "predicted" into "predict"), and merging words within a common message (e.g., "Bank" and "Indonesia" grouped into "Bank Indonesia"). Once this is done, the initial document can be transformed into a document-term matrix, which indicates for each specific text a term's

Figure 3: Topic distributions obtained from text-mining techniques[8]



Source: Hansen (2019)

[7] Such other metrics include, for instance, precision, recall, and F1-score. For binary (two-class) classification, precision is defined as the percentage of times an algorithm makes a correct prediction for the positive class; recall is defined as the percentage of positive class that the algorithm discovers from a given dataset; and the F1-score is the harmonic average of precision and recall.

[8] Distributions obtained from LDA (black, solid line) and EPU dictionary-based index (BBD; red, dashed line). The word-clouds represent word distributions within each topic, with more frequent words shown in larger fonts.

degree of appearance (or non-appearance). This vectoral text representation is made of numerical values that can then be analyzed by quantitative algorithms; for example, to measure the degree of similarity between documents by comparing the related matrixes (Figure 3).

One popular algorithm for working on textual information is the "Latent Dirichlet Allocation" (LDA) [Blei et al. (2003)]. This assumes that documents are distributed by topics, which in turn are distributed by keywords. For example, one document may combine, for a respective 20% and 80%, a "monetary" and an "employment" topic, based on the number of words reflecting this topic distribution (i.e., 20% of them related to words such as "inflation" or "interest rate", and the remaining 80% related to words such as "jobs" and "labor"). Based on these calculations, one can build an indicator measuring how frequently a specific topic appears over time, for instance, to gauge the frequency of the messages related to "recession" – providing useful insights when monitoring the state of the economy.

Besides quantitative algorithms, simpler "dictionary-based methods" can be also employed for analyzing text data. A set of keywords can be selected that are relevant to the topic of interest – for example, a keyword related to "business confidence". Then an index can be constructed based on how frequently these selected keywords appear in a given document, allowing the subject indicator to be assessed (e.g., the evolution of business sentiment) [Tetlock (2007) and Loughran and McDonald (2011)]. A prominent example is the "economic policy uncertainty" (EPU), which quantifies the degree of uncertainty based on the appearance of a set of economic-, policy-, and uncertainty-related keywords in news articles; by the end of 2018, more than 20 country-specific EPU indexes had been compiled [see Baker et al. (2016) and www.policyuncertainty.com/].

2.4 Network analysis

A third important area of big data analytics refers to financial network analysis, which can be seen as the analysis of the relationships between the elements constituting the financial system. Insights into the functioning of this "network" are derived from graphical techniques and representations. This approach can be used to measure how data is connected to other data, clarify how these connections matter, and show how complex systems move in time. It can be particularly effective for big datasets, allowing for the description of complex systems characterized by rich interactions between their components.

The main "modes of analysis" comprise top-down approaches (e.g., analysis of system-wide risk), bottom-up analyses (e.g., analysis of connections between specific nodes of the system), network features analyses (e.g., transmission channels), and agent-based mode ling (e.g., analysis of specific agents involved in the network; for instance, the role of central counterparties (CCPs) in the financial system). Typically, the work will involve three phases, i.e., analysis (data visualization and identification of potential risks), monitoring (e.g., detection of anomalies in real time), and simulation (e.g., scenarios and stress-tests).

In practice, a network is made of elements (nodes), linked to each other either directly or indirectly, and this can be represented by several types of graphs. An important concept is "centrality", which relates to the importance of nodes (or links) in the network, and which can be measured by specific metrics. Another is "community detection", which aims at simplifying the visualization of a large and complex network by regrouping nodes in clusters and filtering noise, through the use of specific machine learning algorithms (see above).

This sort of analysis appears particularly well suited to representing interconnectedness within a system, for instance, by mapping the global value chain across countries and sectors or the types of exposure incurred by financial institutions.[9] One example is the recent work to assess the role of CCPs in the financial system by looking at the connections between them as well as with other financial institutions, such as banking groups, in particular, by considering subsidiary-parent relationships [CPMI-IOSCO (2018)]. This can help to reveal how a disruption originating in one single CCP would affect that CCP's clearing members and, in turn, other CCPs.

3. OPPORTUNITIES FOR CENTRAL BANKS

Big data can play an important role in improving the quality of economic analysis and research, as increasingly recognized by policymakers [Hammer et al. (2017)]. For their part, many central banks are now working on how to make use of the characteristics of financial big datasets in pursuing their mandates [Cœuré (2017)]. Indeed, big data has many advantages in terms of details, flexibility, timeliness, and efficiency, as summarized in the list of their so-called "Vs" – e.g., volumes, variety, velocity, veracity, and value [Laney (2001) and Nymand-Andersen (2016)]. Central banks are interested in developing various pilot projects to better

[9] For a recent example of the monitoring of network effects for global systemic institutions in the context of the DGI, see FSB (2011) and Tissot and Bese Goksu (2018).

Table 1: Relative advantages of designed versus organic data

	DESIGNED DATA	ORGANIC DATA
STRUCTURE	Geographic and socio-economic	Behavior
REPRESENTATIVE	Yes	No
SAMPLE SELECTION	Response rates deteriorating	Extreme
INTRUSIVE	Extremely intrusive	Non-intrusive
COST	Large	Small
CURATION	Well studied	Unclear
PRIVACY	Well protected	Large violations of privacy

Source: Rigobon (2018)

understand the new datasets and techniques, assess their value added in comparison with traditional approaches, and develop concrete "use-cases" [IFC (2015)].

This experience has highlighted the opportunities that big data analytics can provide in key areas of interest to central banks, namely (i) the production of statistical information, (ii) macroeconomic analysis and forecasting, (iii) financial market monitoring, and (iv) financial risk assessment.

3.1 More and enhanced statistical information

Big data can be a useful means of improving the official statistical apparatus. First, it can be an innovative source of support for the current production of official statistics, offering access to a wider set of data, in particular to those that are available in an "organic" way. Unlike statistical surveys and censuses, these data are usually not collected (designed) for a specific statistical purpose, being the by-product of other activities [Groves (2011)]. Their range is quite large, covering transaction data (e.g., prices recorded online), aspirational data (e.g., social media posts, product reviews displayed on the internet), but also various commercial, financial, and administrative indicators. In addition, they present a number of advantages for statistical compilers, such as their rapid availability and the relative ease of collection and processing with modern computing techniques (Table 1) – always noting, however, that actual access to such sources, private or public, can be restricted by commercial and/or confidentiality considerations.

Organic data can be used to enhance existing statistical exercises, especially in improving coverage when this is incomplete. In some advanced economies, the direct web-scraping[10] of online retailers' prices data can, for instance, be used to better measure some specific components of inflation, such as fresh food prices.[11] At the extreme, these data can replace traditional indicators in countries where the official statistical system is underdeveloped. One famous example is the Billion Prices Project [see www.thebillionpricesproject.com, and Cavallo and Rigobon (2016)], which allows inflation indices to be constructed for countries that lack an official and/or comprehensive index. Similarly, a number of central banks in emerging market economies have compiled quick price estimates for selected goods and properties, by directly scraping the information displayed on the web, instead of setting up specific surveys that can be quite time- and resource-intensive.

Second, big data can support a timelier publication of official data, by bridging the time lags before these statistics become available. In particular, the information generated instantaneously by the wide range of web and electronic devices – e.g., search queries – provides high-frequency indicators that can help current economic developments to be tracked more promptly (i.e., through the compilation of advance estimates). Indeed, another objective of the Billion Prices Project is to provide advance information on inflation in a large number of countries, including advanced economies, and with greater frequency – e.g., daily instead of monthly, as with a consumer price index (CPI). Turning to the real economy side, the real time evolution of some "hard" indicators, such as GDP, can now be estimated in advance (nowcast) by using web-based information combined with machine learning algorithms – see Richardson et al. (2019) in the case of New Zealand. The high velocity of big data sources helps to provide more timely information, which can be particularly important during a crisis.

A third benefit is to provide new types of statistics that complement "traditional" statistical datasets. Two important

[10] Web-scraping can be defined as the automated capturing of online information.
[11] Hill (2018) reports that 15% of the US CPI is now collected online.

developments should be noted here. One is the increased availability of digitalized textual information, which allows for the measurement of "soft" indicators such as economic agents' sentiment and expectations – derived, for instance, from social media posts. Traditional statistical surveys can also provide this kind of information, but they typically focus on specific items, e.g., firms' production expectations and consumer confidence [Tissot (2019b)]. In contrast, internet-based sources can cover a much wider range of topics. In addition, they are less intrusive than face-to-face surveys, and may, therefore, better reflect true behaviors and thoughts. A second important element has been the increased use of large granular datasets to improve the compilation of macroeconomic aggregates, allowing for a better understanding of their dispersion [IFC (2016a)] – this type of distribution information is generally missing in the System of National Accounts [SNA; European Commission et al. (2009)] framework.[12]

3.2 Macroeconomic forecasting with big data

Many central banks are already using big datasets for macroeconomic forecasting. Indeed, nowcasting applications as described above can be seen as a specific type of forecasting exercise. For instance, Google Trends data can be used to compile short-term projections of estimates of car sales in the euro area, with a lead time of several weeks over actual publication dates [Nymand-Andersen and Pantelidis (2018)]. Similarly, Gil et al. (2019) argue that big data allows a wider range of indicators to be used for forecasting headline indicators in Spain – for instance Google Trends,[13] uncertainty measures such as the EPU index (see above), or credit card operations, as well as more traditional indicators. The devil is in the details, though, and statisticians need to try several approaches. For instance, some indicators may work well in nowcasting GDP (i.e., its growth rate over the current quarter) but less so in forecasting its future evolution (say, GDP growth one year ahead). Another point is that the internet is not the sole source of indicators that can be used in this context; in fact, some web-based indicators may work less well in nowcasting/forecasting exercises than do traditional business confidence surveys.[14]

In view of these caveats, and considering the vast amount of data potentially available, it may be useful to follow a structured process when conducting such exercises. Sawaengsuksant (2019) recommends a systematic approach when selecting the indicators of interest, such as internet search queries. For instance, key words in Google Trends data could be selected if they satisfied several criteria, depending on their degree of generality, their popularity (i.e., number of searches recorded), their robustness (i.e., sensitivity to small semantic changes), their predictive value (i.e., correlation with macro indicators), and whether the relationship being tested makes sense from an economic perspective.

3.3 Financial market forecasting and monitoring

As in the macroeconomic arena, big data analytics have also proved useful in monitoring and forecasting financial market developments, a key area for central banks. A number of projects in this area facilitate the processing of huge volume of quantitative information in large financial datasets. For instance, Fong and Wu (2019) show that returns in a number of emerging sovereign bond markets can be predicted using various technical trading rules and machine learning techniques to assess their robustness as well as the relative contributions of specific foreign (e.g., U.S. monetary policy) and domestic factors.

Other types of project are looking at less structured data. As an example, Zulen and Wibisono (2019) describe how a text-mining algorithm could be used to measure public expectations for the direction of interest rates in Indonesia. Specifically, a classification algorithm is trained to predict whether a given piece of text indicates an expectation for the future tightening, loosening, or stability of the central bank policy rate. All the newspaper articles discussing potential developments in the policy rate from two weeks prior to monthly policy meetings are collected, and an index of policy rate expectation is produced. This index has facilitated the analysis of the formation of policy rate expectations, usefully complementing other sources (e.g., Bloomberg surveys of market participants). Other types of textual information, such as social media posts and official public statements, could also be usefully considered.

Experience reported by several central banks shows that new big data sources can also help to elucidate developments in financial markets, and shed light on their potential future

[12] Indeed, the SNA highlights the importance of considering the skewed distribution of income and wealth across households but recognizes that getting this information is "not straightforward and not a standard part of the SNA" (2008 SNA, #24.69). It also emphasizes that "there would be considerable analytical advantages in having microdatabases that are fully compatible with the corresponding macroeconomic accounts" (2008 SNA, #1.59). An important recommendation of the second phase of the DGI aims at addressing these issues [FSB-IMF (2015)].

[13] See https://trends.google.com/trends/. Google Trends provide indexes of the number of Google searches of given keywords. The indexes can be further segregated based on countries and provinces.

[14] For the use of nowcasting in forecasting "bridge models" using traditional statistics and confidence surveys, see Carnot et al. (2011).

direction. Sakiyama and Kobayashi (2018) have used high-frequency "tick" transaction data to assess market liquidity in the Japanese government bond market, and hence the risk of potential abrupt price changes. Similarly, the Bank of England has set up specific projects to identify forex market dynamics and liquidity at times of large market movements – e.g., when the Swiss National Bank decided to remove the EUR/CHF floor in January 2015 [Cielinska et al. (2017)].

3.4 Financial risk assessment

Big data sources and techniques can also facilitate financial risk assessment and surveillance exercises that sit at the core of central banks' mandates – for both those in charge of micro-financial supervision and those focusing mainly on financial stability issues and macro-financial supervision [Tissot (2019a)]. In particular, the development of big data analytics has opened promising avenues for using the vast amounts of information entailed in granular datasets to assess financial risks.

To start with, they allow new types of indicators to be derived, as highlighted by Petropoulos et al. (2019) for the analysis of the financial strength of individual firms in Greece. Based on the granular information collected in the central bank's supervisory database (covering around 200,000 borrowers over a decade), a deep learning technique with a specific classification algorithm[15] was used to forecast the likelihood of default for each loan outstanding. To facilitate policy monitoring work, this approach was complemented by a dimensionality reduction algorithm to reduce the number of variables to be considered.

Moreover, big data analytics can help to enhance existing financial sector assessment processes (e.g., regtech), by extending conventional methodologies and providing additional insights – in terms of, for example, financial sentiment analysis, early warning systems, stress-test exercises, and network analysis. For instance, with the use of network analytics for systemic risk measurement and contagion effects [Langfield and Soramäki (2016)], the application of text analysis techniques to corporate e-mails and news for risk assessment [Das et al. (2019)], the assistance of complex visualization techniques to support data exploration and monitoring for large-scale financial networks [Heijmans et al. (2016)], etc. Yet, these approaches underline the importance of having a sound theoretical framework to interpret the signals provided by disparate sources as well as to detect unusual, odd patterns in the data. They also highlight the important role played by model simplicity and transparency, the benefit of a multidisciplinary approach, and the high IT and staff costs involved.

4. THE USE OF BIG DATA IN CRAFTING CENTRAL BANK POLICY: ORGANIZATION AND CHALLENGES

Central bank experience suggests that the opportunities provided by big data sources and related analytical techniques can be significant, supporting a wide range of areas of policy interest. But how should central banks organize themselves to make the most of these opportunities? And what are the key challenges?

4.1 Organizational issues

Central banks' tasks cover a wide range of topics that can greatly benefit from big data. For instance, central bankers need near-real-time and higher-frequency snapshots of the macroeconomy's state, its potential evolution (central scenario), and the risks associated with this outlook (e.g., early warning indicators and assessment of turning points). At the same time, their interest in financial stability issues calls for the ability to zoom in and get insights at the micro-level – see the ongoing initiatives among European central banks to develop very granular datasets on security-by-security issuance and holdings as well as on loan-by-loan transactions [the AnaCredit project; Schubert (2016)].

This puts a premium on information systems that can support this diversity of approaches. One reported lesson of the data lake platform project being developed at the French central bank [Villeroy de Galhau (2017)] is that a multidisciplinary and granular data platform is required to supply flexible and innovative services to a wide range of internal users. The aim is to provide key data management services to support multiple activities, covering data collection, supply (access), quality management, storage, sharing, analytics, and dissemination. From a similar perspective, the Deutsche Bundesbank has set up an integrated microdata-based information and analysis system (IMIDIAS) to facilitate the handling of granular data used to support its activities [Staab (2017)]. It has also worked on fostering internal as well as external research on this information to gain new insights and facilitate policy analysis. Moreover, the Bundesbank actively supports the International

[15] eXtreme Gradient Boosting (XGBoost), which is commonly used in decision tree-based algorithms; see Chen and Guestrin (2016) and https://xgboost.readthedocs.io/en/latest/.

Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA) [Bender et al. (2018)].

A key takeaway is that the development of an adequate information system is only one element of a more comprehensive strategy to make the most of big data at central banks. This usually requires the use of various techniques, e.g., machine learning, text-mining, natural language processing, and visualization techniques; and it also has to be backed by an extensive staff training program on data analytics. As an example, several use-cases involving data science have been developed at the Netherlands Bank – e.g., in the areas of credit risk, contagion risk, CCP risk, and stress-testing in specific market segments. An important outcome has been the recognition of the important role played not only by the techniques used but also by the staff, organization, and culture.

4.2 Challenges and limitations

In practice, important challenges remain, especially in handling and using big data sources and techniques.

Handling big data can be resource-intensive, especially in collecting and accessing the information, which can require new, expensive IT equipment, as well as state-of-the-art data security. Staff costs should not be underestimated too, as suggested by the experience reported for many central banks. First, large micro-datasets on financial transactions often have to be corrected for false attributes, missing points, outliers, etc. [Cagala (2017)]; this cleaning work may often require the bulk of the time of the statisticians working with these data. Second, a much wider set of profiles – e.g., statisticians, IT specialists, data scientists, and also lawyers – are needed to work in big data multidisciplinary teams; ensuring a balanced skillset and working culture may be challenging. Third, there is a risk of a “war for talent” when attracting the right candidates, especially vis-à-vis private sector firms that are heavily investing in big data; public compensation and career systems may be less than ideally calibrated for this competition. In addition, and as seen above, a key organizational consideration is how to integrate the data collected into a coherent and comprehensive information model. The challenge for central bank statisticians is thus to make the best use of available data that were not originally designed for specific statistical purposes and can be overwhelming (with the risk of too much “fat data”, and too little valuable information). In most cases, this requires significant preparatory work and sound data governance principles, covering data quality management processes (e.g., deletion of redundant information), the setup of adequate documentation (e.g., metadata), and the allocation of clear responsibilities (e.g., “who does what for what purpose”) and controls.

Using big data is also challenging for public authorities. One key limitation relates to the underlying quality of the information as noted above. This challenge can be reinforced by the large variety of big data formats, especially when the information collected is not well structured. Moreover, big data analytics rely frequently on correlation analysis, which can reflect coincidence as well as causality patterns. Furthermore, the veracity of the information collected may prove insufficient. Big datasets may often cover entire populations, so by construction there is little sampling error to correct for, unlike with traditional statistical surveys. But a common public misperception is that, because big datasets are extremely large, they are automatically representative of the true population of interest. Yet, this is not guaranteed, and in fact the composition bias can be quite significant, in particular as compared with much smaller traditional probabilistic samples [Meng (2014)]. For example, when measuring prices online, one must realize that not all transactions are conducted on the internet. The measurement bias can be problematic if online prices are significantly different from those observed in physical stores, or if the products consumers buy online are different to those they buy offline.

These challenges are reinforced by two distinctive features of central banks – the first being their independence and the importance they accord to preserving public trust. Since the quality of big datasets may not be at the standard required for official statistics, “misusing” them as the basis of policy actions could raise ethical, reputational, as well as efficiency issues. Similarly, if the confidentiality of the data analyzed is not carefully protected, this could undermine public confidence, in turn calling into question the authorities’ competence in collecting, processing, and disseminating information derived from big data, as well as in taking policy decisions inferred from such data. This implies that central banks would generally seek to provide reassurances that data are used only for appropriate reasons, that only a limited number of staff can access them, and that they are stored securely. The ongoing push to access more detailed data (often down to individual transaction level) reinforces the need for careful consideration of the requirement to safeguard the privacy of the individuals and firms involved.

A second feature is that central banks are policymaking institutions whose actions are influencing the financial system and thereby the information collected on it. Hence, there is a feedback loop between the financial big data collected, its use for designing policy measures, and the actions taken by market participants in response. As a result, any move to measure a phenomenon can lead to a change in the underlying reality, underscoring the relevance of the famous Lucas critique for policymakers [Lucas (1976)].

5. CONCLUSION

5.1 Main lessons

Central banks' various experiences in using big data analytics and associated AI techniques have highlighted the following points:

- Big data offers new types of data source that complement more traditional varieties of statistics. These sources include Google searches, real estate and consumer prices displayed on the internet, and indicators of economic agents' sentiments and expectations (e.g., social media).
- Thanks to IT innovation, new techniques can be used to collect data (e.g., web-scraping), process textual information (text-mining), match different data sources (e.g., fuzzy-matching), extract relevant information (e.g., machine learning), and communicate or display pertinent indicators (e.g., interactive dashboards).
- In particular, big data techniques, such as decision trees, may shed interesting light on the decision-making processes of economic agents, e.g., how investors behave in financial markets. As another example, indicators of economic uncertainty extracted from news articles could help explain movements of macroeconomic indicators. This illustrates big data's potential in providing insights not only into what happened, but also into what might happen and why.
- In turn, these new insights can usefully support central bank policies in a wide range of areas, such as market information (e.g., credit risk analysis), economic forecasting (e.g., nowcasting), financial stability assessments (e.g., network analysis), and external communications (e.g., measurement of agents' perceptions). Interestingly, the approach can be very granular, helping to target specific markets, institutions, instruments, and locations (e.g., zip codes) and, in particular, to support macroprudential policies. Moreover, big data indicators are often more timely than "traditional" statistics – for instance, labor indicators can be extracted from online job advertisements almost in real time.
- As a note of caution, feedback from central bank pilot projects consistently highlights the complex privacy implications of dealing with big data, and the associated reputational risks. Moreover, while big data applications, such as machine learning algorithms, can excel in terms

of predictive performance, they can lend themselves more to explaining what is happening rather than why. As such, they may be exposed to public criticism when insights gained in this way are used to justify policy decisions.

- Another concern is that, as big data samples are often far from representative (e.g., not everyone is on Facebook, and even fewer are on Twitter), they may not be as reliable as they seem. Lastly, there is a risk that collecting and processing big data will be hindered by privacy laws and/or changes in market participants. Relevant authorities should coordinate their efforts so that they can utilize the advantages of big data analytics without compromising data privacy and confidentiality.

5.2 Looking forward

Taking stock of the implementation of big data projects in the central bank community shows that new big data-related sources of information and analytical techniques can provide various benefits for policymakers. Yet, big data is still seen as complementing, rather than replacing, present statistical frameworks. It raises a number of difficult challenges, not least in terms of accuracy, transparency, confidentiality, and ethical considerations. These limitations apply to big data sources as well as to the techniques that are being developed for their analysis. In particular, one major drawback of big data analytics is their black-box character, a difficulty reinforced by the frequent use of fancy names even for simple things (buzzwords). This can be a challenge for policymakers who need to communicate the rationale behind their analyses and decisions as transparently as possible. Moreover, important uncertainties remain on a number of aspects related to information technology and infrastructures, such as the potential use of cloud-based services and the development of new processes (e.g., cryptography, anonymization techniques) to facilitate the use of micro-level data without compromising confidentiality.

One important point when discussing these issues is that central banks do not work in isolation. They need to explain to the general public how the new data can be used for crafting better policies, say, by providing new insights into the functioning of the financial system, clarifying its changing structure, improving policy design, and evaluating the result of policy actions [Bholat (2015)]. But they also need to transparently recognize the associated risks, and to clearly state the safeguards provided in terms of confidentiality protection, access rights, and data governance. Ideally, if big data is to be used for policymaking, the same quality of standards and frameworks that relate to traditional official statistics should be applied, such as transparency of sources, methodology, reliability, and consistency over time. This will be key to facilitating a greater use of this new information as well as its effective sharing between public bodies.[16]

> “*As both data users and data producers, central banks are in an ideal position to ensure that big data can be transformed into useful information in support of policy.*”

Looking forward, it is still unclear whether and how far big data developments will trigger a change in the business models of central banks, given that they are relatively new to exploiting this type of information and techniques. Central banks have historically focused more on analyzing data and less on compiling them. They are now increasingly engaged in statistical production, reflecting the data collections initiated after the GFC as well as the growing importance of financial channels in economic activities – see, for instance, the substantial involvement of central banks in the compilation of financial accounts, a key element of the SNA framework [van de Ven and Fano (2017)]. As both data users and data producers, they are, therefore, in an ideal position to ensure that big data can be transformed into useful information in support of policy.

[16] On the general data-sharing issues faced by central banks, see IFC (2016b).

REFERENCES

Baker, S., N. Bloom, and S. Davis, 2016, "Measuring economic policy uncertainty," Quarterly Journal of Economics 131:4, 1593–1636

Bean, C., 2016, Independent review of UK economic statistics, March

Bender, S., C. Hirsch, R. Kirchner, O. Bover, M. Ortega, G. D'Alessio, L. Teles Dias, P. Guimarães, R. Lacroix, M. Lyon, and E. Witt, 2016, "INEXDA – the Granular Data Network," IFC Working Papers no. 18, October

Bholat, D., 2015, "Big data and central banks," Bank of England, Quarterly Bulletin, March

Blei, D., A. Ng, and M. Jordan, 2003, "Latent dirichlet allocation," Journal of Machine Learning Research 3, 993–1022

Cagala, T., 2017, "Improving data quality and closing data gaps with machine learning," IFC Bulletin no. 46, December

Carnot, N., V. Koen, and B. Tissot, 2011, Economic Forecasting and Policy, second edition, Palgrave Macmillan

Cavallo, A., and R. Rigobon, 2016, "The billion prices project: using online prices for measurement and research," Journal of Economic Perspectives 30:2, 151–178

Chen, T., and C. Guestrin, 2016, "Xgboost: a scalable tree boosting system," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794

Cielinska, O., A. Joseph, U. Shreyas, J. Tanner, and M. Vasios, 2017, "Gauging market dynamics using trade repository data: the case of the Swiss franc de-pegging," Bank of England, Financial Stability Papers no. 41, January

Coeuré, B., 2017, "Policy analysis with big data", speech at the conference on "Economic and financial regulation in the era of big data," Bank of France, Paris, November

CPMI- IOSCO, 2018, "Framework for supervisory stress testing of central counterparties (CCPs)," Committee on Payments and Market Infrastructures (CPMI) and Board of the International Organization of Securities Commissions (IOSCO), April

Das, S., S. Kim, and B. Kothari, 2019, "Zero-revelation RegTech: Detecting risk through linguistic analysis of corporate emails and news," Journal of Financial Data Science 2, 8-34

Economist, 2017, "The world's most valuable resource is no longer oil, but data," May 6

E.C., IMF, OECD, U.N., and World Bank, 2009, "System of National Accounts 2008," European Commission, International Monetary Fund, Organisation for Economic Cooperation and Development, United Nations, and World Bank

FSB, 2011, "Understanding financial linkages: a common data template for global systemically important banks," FSB Consultation Papers, Financial Stability Board

FSB, 2017, "Artificial intelligence and machine learning in financial services – market developments and financial stability implications," Financial Stability Board, November

FSB-IMF, 2009, "The financial crisis and information gaps," Financial Stability Board and International Monetary Fund

FSB-IMF, 2015, "The financial crisis and information gaps – Sixth Implementation Progress Report of the G20 Data Gaps Initiative," Financial Stability Board and International Monetary Fund

Fong, T., and G. Wu, 2019, "Predictability in sovereign bond returns using technical trading rule: do developed and emerging markets differ?," IFC Bulletin no. 50, May

Gil, M., J. J Pérez, A. J Sánchez, and A. Urtasun, 2019, "Nowcasting private consumption: traditional indicators, uncertainty measures, credit cards and some internet data," IFC Bulletin no. 50, May

Groves, R., 2011, "Designed data and organic data," in the Director's Blog of the U.S. Census Bureau, https://bit.ly/2kJLscq

Haldane, A., 2018, "Will big data keep its promise?" speech at the Bank of England Data Analytics for Finance and Macro Research Centre, King's Business School, April 19

Hammer, C., D. Kostroch, G. Quirós, and staff of the IMF Statistics Department (STA) Internal Group, 2017, "Big data: potential, challenges, and statistical implications," IMF Staff Discussion Notes no. 17/06, September

Hansen, S., 2019, "Introduction to text mining," IFC Bulletin no. 50, May

Heijmans, R., R. Heuver, C. Levallois, and I. van Lelyveld, 2016, "Dynamic visualization of large financial networks," Journal of Network Theory in Finance 2:2, 57–79

Hill, S., 2018, "The big data revolution in economic statistics: waiting for Godot ... and government funding," Goldman Sachs US Economics Analyst, 6 May

IFC, 2015, "Central banks' use of and interest in 'big data'," Irving Fisher Committee on Central Bank Statistics, IFC Report, October

IFC, 2016a, "Combining micro and macro statistical data for financial stability analysis," Irving Fisher Committee on Central Bank Statistics, IFC Bulletin no. 41, May

IFC, 2016b, "The sharing of micro data – a central bank perspective," Irving Fisher Committee on Central Bank Statistics, IFC Report, December

IFC, 2017a, "Big data," Irving Fisher Committee on Central Bank Statistics, IFC Bulletin no. 44, September

IFC, 2017b, "Uses of central balance sheet data offices' information," Irving Fisher Committee on Central Bank Statistics, IFC Bulletin no. 45, October

IFC, 2018, "Central banks and trade repositories derivatives data," Irving Fisher Committee on Central Bank Statistics, IFC Report, October

IFC, 2019, "The use of big data analytics and artificial intelligence in central banking," Irving Fisher Committee on Central Bank Statistics, IFC Bulletin no. 50, May

Laney, D., 2001, "3D data management: controlling data volume, velocity, and variety," META Group (now Gartner)

Langfield, S., and K. Soramäki, 2016, "Interbank exposure networks," Computational Economics 47:1, 3-17

Loughran, T., and B. McDonald, 2011, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," Journal of Finance 66:1, 35–65

Lucas, R., 1976, "Econometric policy evaluation: A critique," Carnegie-Rochester Conference Series on Public Policy 1:1, 19–46

Meeting of the Expert Group on International Statistical Classifications, 2015, "Classification of types of big data," United Nations Department of Economic and Social Affairs, ESA/STAT/AC.289/26, May

Meng, X., 2014, "A trio of inference problems that could win you a Nobel Prize in statistics (if you help fund it)," in Lin, X., C. Genest, D. Banks, G. Molenberghs, D. Scott, and J.-L. Wang (eds), Past, present, and future of statistical science, Chapman and Hall, 537–562

Nymand-Andersen, P., 2016, "Big data – the hunt for timely insights and decision certainty: central banking reflections on the use of big data for policy purposes," IFC Working Papers no. 14, February

Nymand-Andersen, P., and E. Pantelidis, 2018, "Google econometrics: nowcasting euro area car sales and big data quality requirements," European Central Bank, Statistics Paper Series no. 30, November

Petropoulos, A., V. Siakoulis, E. Stravroulakis, and A. Klamargias, 2019, "A robust machine learning approach for credit risk analysis of large loan-level datasets using deep learning and extreme gradient boosting," IFC Bulletin no. 50, May

Richardson, A., T. van Florenstein Mulder, and T. Vehbi, 2019, "Nowcasting New Zealand GDP using machine learning algorithms," IFC Bulletin no. 50, May

Rigobon, R., 2018, “Promise: measuring from inflation to discrimination,” presentation given at the workshop on “Big data for central bank policies,” Bank Indonesia, Bali, 23–25 July

Sakiyama, T., and S. Kobayashi, 2018, “Liquidity in the JGB cash market: an evaluation from detailed transaction data,” Bank of Japan, Reports & Research Papers, March

Sawaengsuksant, P., 2019, “Standardized approach in developing economic indicators using internet searching applications,” IFC Bulletin no. 50, May

Schubert, A., 2016, “AnaCredit: banking with (pretty) big data,” Central Banking Focus Report

Staab, P., 2017, “The Bundesbank’s house of micro data: Standardization as a success factor enabling data-sharing for analytical and research purposes,” IFC Bulletin no. 43, March

Tetlock, P., 2007, “Giving content to investor sentiment: the role of media in the stock market,” Journal of Finance 62:3, 1139–1168

Tissot, B., and E. Bese Goksu, 2018, “Monitoring systemic institutions for the analysis of micro-macro linkages and network effects,” Journal of Mathematics and Statistical Science 4:4, 129-136

Tissot, B., (2019a), “Making the most of big data for financial stability purposes”, in Strydom, S., and M. Strydom (eds), Big data governance and perspectives in knowledge management, IGI Global, 1–24

Tissot, B., 2019b, “The role of big data and surveys in measuring and predicting inflation,” International Statistical Institute World Statistics Congress, August

Van de Ven, P., and D. Fano, 2017, Understanding financial accounts, OECD Publishing, Paris

Villeroy de Galhau, F., 2017, “Economic and financial regulation in the era of big data,” speech at the Bank of France conference, Paris, November

Zulen, A., and O. Wibisono, 2019, “Measuring stakeholders’ expectations for the central bank’s policy rate,” IFC Bulletin no. 50, May

This document was produced for information purposes only and is for the exclusive use of the recipient.

This publication has been prepared for general guidance purposes, and is indicative and subject to change. It does not constitute professional advice. You should not act upon the information contained in this publication without obtaining specific professional advice. No representation or warranty (whether express or implied) is given as to the accuracy or completeness of the information contained in this publication and The Capital Markets Company BVBA and its affiliated companies globally (collectively "Capco") does not, to the extent permissible by law, assume any liability or duty of care for any consequences of the acts or omissions of those relying on information contained in this publication, or for any decision taken based upon it.

ABOUT CAPCO

Capco is a global technology and management consultancy dedicated to the financial services industry. Our professionals combine innovative thinking with unrivalled industry knowledge to offer our clients consulting expertise, complex technology and package integration, transformation delivery, and managed services, to move their organizations forward.

Through our collaborative and efficient approach, we help our clients successfully innovate, increase revenue, manage risk and regulatory change, reduce costs, and enhance controls. We specialize primarily in banking, capital markets, wealth and asset management and insurance. We also have an energy consulting practice in the US. We serve our clients from offices in leading financial centers across the Americas, Europe, and Asia Pacific.

WORLDWIDE OFFICES

APAC
Bangalore
Bangkok
Hong Kong
Kuala Lumpur
Pune
Singapore

EUROPE
Bratislava
Brussels
Dusseldorf
Edinburgh
Frankfurt
Geneva
London
Paris
Vienna
Warsaw
Zurich

NORTH AMERICA
Charlotte
Chicago
Dallas
Houston
New York
Orlando
Toronto
Tysons Corner
Washington, DC

SOUTH AMERICA
São Paulo

FSC
www.fsc.org
MIX
Paper from
responsible sources
FSC® C001648

WWW.CAPCO.COM

CAPCO