Irving Fisher Committee on Central Bank Statistics

IFC Working Papers

No 21

Big data in Asian central banks

By Giulio Cornelli, Sebastian Doerr, Leonardo Gambacorta and Bruno Tissot

February 2022

BANK FOR INTERNATIONAL SETTLEMENTS

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

ISSN 1991-7511 (online)
ISBN 978-92-9259-533-3 (online)

Big data in Asian central banks

Giulio Cornelli, Sebastian Doerr, Leonardo Gambacorta and Bruno Tissot[1]

Contents

[1] Respectively Senior Financial Market Analyst (Giulio.Cornelli@bis.org); Economist (Sebastian.Doerr@bis.org); Head of Innovation and Digital Economy (Leonardo.Gambacorta@bis.org); Head of Statistics and Research Support, Bank for International Settlements (BIS) and Head of the Secretariat of the Irving Fisher Committee on Central Bank Statistics (IFC) (bruno.tissot@bis.org).

We would like to thank Jose Maria Serena and Fernando Perez Cruz for their advice and input. For comments and suggestions, we also thank Redentor Paolo Alegre Jr, Gianni Amisano, Douglas Araujo, Claudio Borio, Agustin Carstens, Stijn Claessens, Jon Frost, Michel Juillard, Julian Langer, Juri Marcucci, Li Ming, Kuniko Moriya, Luiz Awazu Pereira, Rafael Schmidt, Hyun Song Shin and Helio Vale. The views expressed are those of the authors and not necessarily those of the BIS or the IFC.

Abstract

This paper reviews the use of big data in Asian central banks, leveraging on a survey conducted among the members of the Irving Fisher Committee. The analysis reveals four main insights. First, Asian central banks define big data in a more encompassing way that includes unstructured non-traditional as well as structured data sets. Second, interest in big data appears higher in Asia, including at the senior policy level; the focus is in particular on projects developed to process natural language, conduct nowcasting/monitoring exercises, and develop applications to extract economy insights as well as suptech/regtech solutions. Third, Asian central banks report dealing with big data to support a wide range of tasks. Fourth, big data poses new challenges, with specific attention paid in the region to cyber security and data strategy. As a result, there is a growing need for international policy cooperation, especially among public authorities in Asia to facilitate the use of payments data and promote innovative technological solutions.

Keywords: Asian central banks, artificial intelligence, big data, data science, international cooperation

JEL codes: G17, G18, G23, G32

1. Introduction

Big data sources are developing fast, and applications for making use of this new information are flourishing in parallel. This trend, which is particularly pronounced in Asia, primarily reflects the impact of digitalisation, with the development of the "internet of things" and the ever-increasing ability to digitally process "traditional" information, such as text. It is also a consequence of the large databases that have been created as a by-product of the complex operations taking place in modern societies. Additionally, vast amounts of data have emerged in the administrative, commercial and financial realms, an evolution spurred by the important data collection strategies undertaken after the great financial crisis of 2007–09 to address the information challenges posed by developments in the financial sector. We now live in the "age of big data" (Forbes (2012)).

Central banks are no exception to this general picture (Buch (2019)). They have shown an increasing interest in using big data in recent years, as already documented extensively by the Irving Fisher Committee (IFC) on Central Bank Statistics (IFC (2017), Tissot (2017), Nymand-Andersen (2016), Mehrhoff (2019)). Central bank big data-related work covers a variety of areas, including monetary policy and financial stability as well as research and the production of official statistics. However, in contrast to the rapid pace of innovation seen in the private sector, big data applications supporting central banks' operational work were developed only slowly initially. This tended to reflect a number of constraints, such as a lack of adequate resources as well as the intrinsic challenges associated with using big data sources to support public policy. Yet, in recent years, central banks' use of big data has proliferated, especially among Asian countries.

Will central banks catch up and transform the way they operate to further benefit from the information revolution? Or will their use of big data sources and applications progress only gradually due to the inherent specificities of their mandates and processes? To shed light on these issues, this paper reviews **the use of big data and machine learning in the Asian central bank community, leveraging on a survey conducted in 2020 among the members of the IFC** (IFC (2021a)). To this end, this paper analyses the responses from seven Asian central banks, with a specific focus on their reported big data projects (cf Appendix).[2]

The approach dealt with the following key questions: What constitutes big data for central banks, and how strong is central banks' interest in it? Have central banks been increasing their use of big data and, if so, what were the main applications developed? And finally, which constraints are faced when using big data and how can they be overcome? To address these issues, Asian central banks' answers to the 2021 survey were compared with those of their peers in the rest of the world.

This analysis uncovered **four main insights**.

First, Asian central banks have a **comprehensive view of big data**, which can comprise very different types of data sets. First and foremost, it includes large

[2] The list of Asian central banks includes: Bangko Sentral ng Pilipinas (BSP), Bank Indonesia (BI), Bank of Japan (BoJ), Bank of Thailand (BoT), Bank Negara Malaysia (BNM), Monetary Authority of Macao (MAM), Reserve Bank of India (RBI). Almost two thirds of the 92 IFC institutional members at that time answered the survey. More information on the survey is contained in IFC (2021a) and Doerr et al (2021).

"non-traditional" (or unstructured) data often characterised by high volume, velocity and variety and that must be processed using innovative technologies. Yet for the vast majority (85%) of respondents in Asia, big data also includes large "traditional" (ie structured) data sets. These can be the outcome of explicit reporting requirements set by public regulators; they are also often "organic" by-products collected as a result of commercial (eg payment transactions), financial (eg tick-by-tick price quotes observed in financial markets) and administrative (eg files collected by public institutions) activities – these data are often referred to as "financial big data". In contrast, only 60% of central banks outside Asia do include such traditional data sets in the concept of "big data". Potentially, the relatively large footprint of big techs in Asia has stimulated the discussion in the region (Cornelli et al (2020)).

Second, **interest** in big data is high in Asia: around two thirds (60%) of central banks in the region mentioned that they discuss big data issues extensively, while this is reported to be the case by only a minority (42%) of their counterparts in the rest of the world. Moreover, all Asian central banks in the survey indicated a high to very-high level of interest also at the senior policy level, while this was the case for only 58% of their counterparts in other regions.

Third, and turning to **concrete use cases**, 68% of Asian central banks report dealing with big data to support economic research, monetary and financial stability policies as well as their statistical production tasks. This is comparable to the numbers reported in the rest of the world (64%). The big data projects undertaken in this context typically involve four main types of applications: natural language processing (NLP), nowcasting exercises (including to support their statistical processing tasks), applications to extract information on the state of the economy from granular financial data and other non-traditional sources as well as suptech/regtech applications.

Fourth, the survey shows that Asian central banks discuss extensively **the new challenges** posed by the advent of big data. A major one is setting up a reliable and high-powered IT infrastructure. While many institutions have undertaken important initiatives to develop adequate platforms to facilitate the storage and processing of very large and complex data sets (IFC (2020)), progress has varied in the region. This is in part because of the need to hire and train staff, which is difficult due to the limited supply of candidates with the necessary skills (eg, data scientists). Other challenges include the legal basis for using private data and the safety, ethical and privacy concerns this entails, as well as the "fairness" and accuracy of algorithms trained on preclassified and/or unrepresentative data sets. Data quality and governance issues are also significant, since much of the new big data collected as a by-product of economic or social activities needs to be curated before proper statistical analysis can be conducted (IFC (2021b)). These challenges are generally seen as equally important among different central banks across the world. One notable point is that cyber security and the development of a formal strategy for the use of big data are topics that appear to be higher on the agenda of Asian central banks compared to their counterparts in other regions.

The rest of the paper is organised as follow. Section 2 provides an overview of how Asian central banks define big data. Sections 3 illustrates in which fields they use or plan to use big data and discusses specific use cases. Section 4 reviews the main challenges in the use of machine learning and big data. Section 5 discusses how cooperation among public authorities could relax the constraints on collecting, storing and analysing big data. Section 6 concludes.

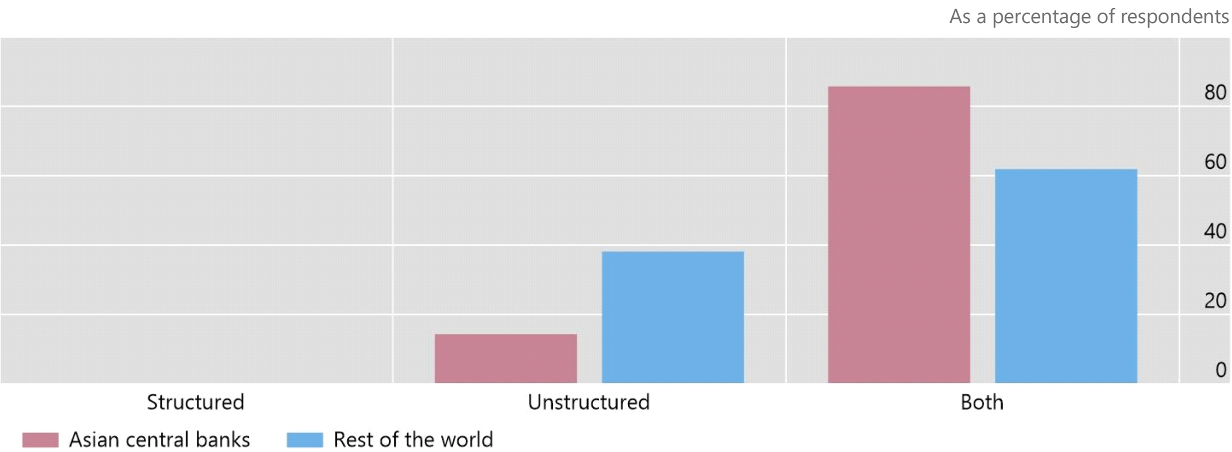
2. What is central banks' definition of big data?

The definition of big data is not unique, as it pertains to the specific angle of its use. In general, big data can be defined in terms of **volume, velocity and variety** (the so-called 3Vs). The reason is that for data to be "big", they must not only have high volume and high velocity, but also come in multiple varieties. Yet there are also many different views on what defines "big data".[3]

In **practice**, big data can include the information generated from a wide variety of sources, such as social media, web-based activities, machine sensors, or financial, administrative or business operations. This comprehensive view of big data is confirmed by the survey results for Asian central banks. Certainly, no central bank considers traditional data alone as big data. But as reported in Graph 1, only 14% of the respondents define big data exclusively as large non-traditional or unstructured data that require new techniques for the analysis (in contrast, almost 40% of their counterparts in the rest of the world have such a narrow definition). The remaining 86% of Asian respondents also include traditional and structured data sets in their definition of big data. These structured data sets comprise those collected for administrative or regulatory/supervisory purposes, often labelled as "financial big data" (Cœuré (2017), Draghi (2018)).

Based on the results from the survey, a **comprehensive definition of big data** would therefore cover all types of data sets that require non-standard technologies to be analysed. The reason for this is, in part, that traditional statistical techniques face hurdles when applied to unstructured data. For instance, to analyse handwritten text, it must first be turned into structured data, as is done for instance with NLP algorithms.

Central bank definitions of big data and main sources | Graph 1



The sample includes 7 Asian central banks and 43 non-Asian central banks. Respondents could select multiple options. Non-traditional data include "unstructured data sets that require new tools to clean and prepare", "data sets with a large number of observations in the time series", "data sets that have not been part of your traditional pool", and "data sets with a large number of observations in the cross-section".

Sources: IFC (2021a); authors' calculations.

3 Occasionally, veracity is also added, as big data is often collected from open sources; moreover, the literature is quite diverse and can refer to a much larger number of "Vs" (Tissot (2019)).

There is a **variety of raw data sources** used by Asian central banks for analysis. These range from structured administrative data sets such as credit registries to non-traditional data obtained from newspapers and online portals or by scraping the web. This type of information – including the data produced by the internet itself – may not necessarily be "big", but it is complex and cannot be easily analysed with traditional statistical techniques tailored to numerical data sets. Instead, it requires specific tools to be cleaned and properly prepared. However, in some instances it is possible to acquire these data from private providers in an already aggregated and organised form.

Three examples are worth mentioning. First, mobility reports, which provide aggregate commuting trends obtained through GPS from mobile phones and which were able to support the monitoring of households' access to recreation areas when the Covid-19 pandemic struck in 2020 (see Bank of Japan (2020)). The second example relates to internet searches, such as Google Trends, that can be used to assess developments in real time – for instance, expectations on developments in the labor market (Doerr and Gambacorta (2020a,b)) or car sales (Nymand-Andersen and Pantelidis (2018)). A third source of unstructured information for central banks is text in printed format, such as newspaper articles, firms' financial statements, official press releases, etc.

While central banks have substantial experience with large, structured data sets, typically of a financial nature, they have only recently started to explore unstructured data. As discussed above, the analysis of unstructured data requires the application of specific tools. They are often the by-product of corporate or consumer activity and before they are analysed, they must be cleaned and curated, ie organised and integrated into existing structures.

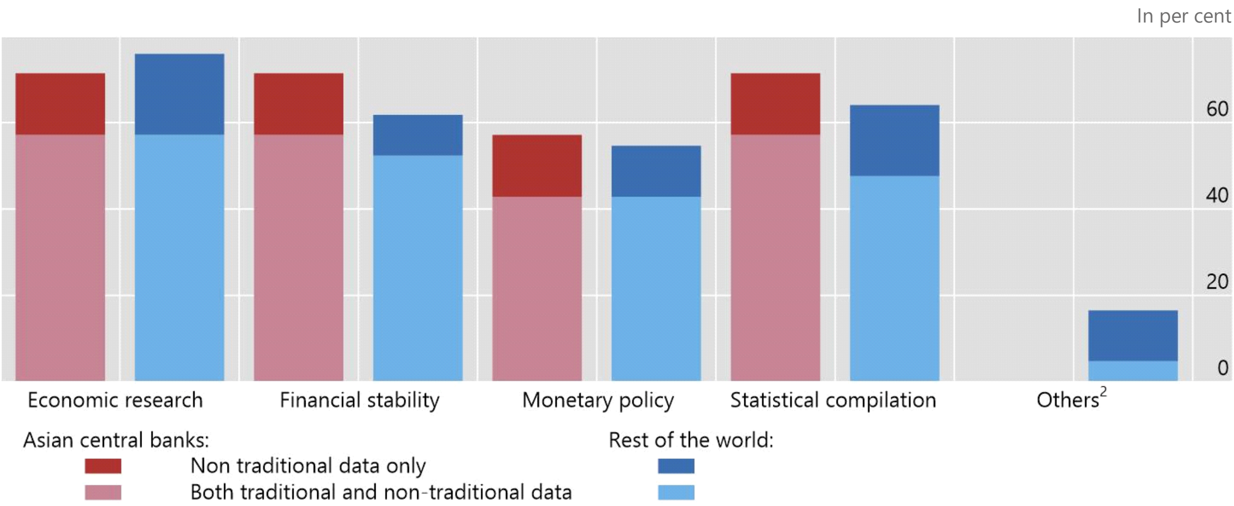
3. How do Asian central banks use big data?

According to the 2020 IFC survey, central banks and supervisory authorities are **rapidly adopting big data and machine learning**: the share of central banks currently using big data has risen to 80% globally, up from just 30% in 2015. This share has risen from 33% to 86% when looking specifically into Asia. Moreover, around 60% of central banks in the region reported that they discuss big data issues extensively, a ratio that is significantly above the one (42%) observed in the rest of the world. Furthermore, all Asian respondents indicated a high to very-high level of interest at the senior policy level, compared to only 58% outside the region.

Big data is used in a **variety of areas**, including research as well as monetary policy and financial stability. Asian central banks (represented by the red bars in Graph 2) appear to use big data in most areas by more than their peers (blue bars), except for research purposes. In particular, they process non-traditional data (darker bars) to a greater extent to support monetary and financial stability policies – including for specific supervisory and regulatory purposes (suptech and regtech).

Purposes for which central banks use big data

Graph 2



[1] Respondents could select multiple options. See footnote to Graph 1 for details on how institutions define big data and on the sample composition. [2] Includes "monitoring crypto assets", "cyber security" and "network analysis".

Sources: IFC (2021a); authors' calculations.

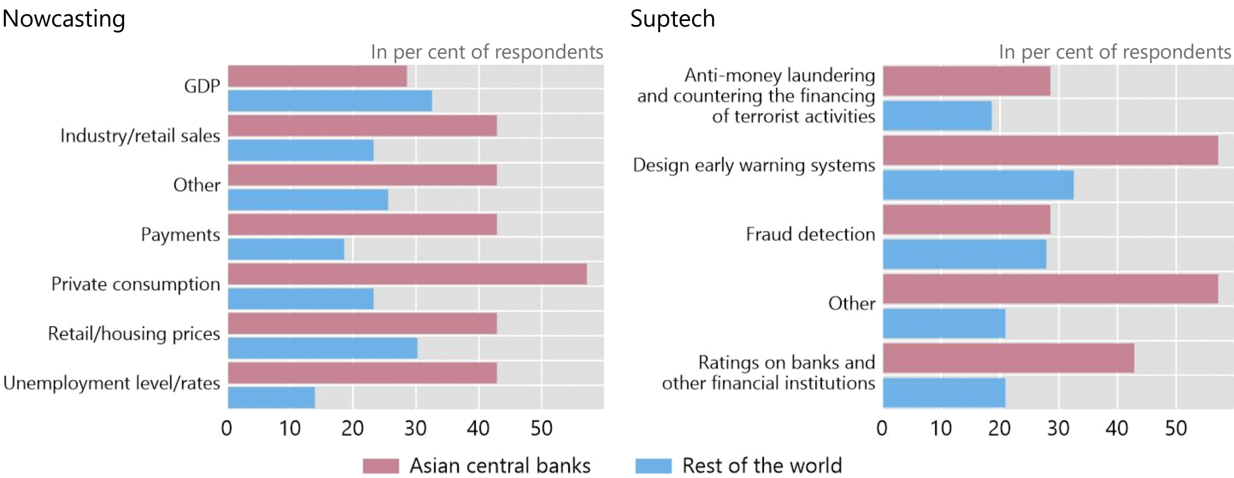
The big data projects undertaken by Asian central banks involve **four main types of applications**: NLP, nowcasting exercises, applications to extract economy wide insight from granular financial data and other non-traditional sources, and suptech/regtech applications. A list of selected big data projects in Asian central banks is provided in the Appendix.

A first type of application uses **textual information through NLP**. The goal is generally to turn qualitative text-based intelligence into numerical format. One example has been the computation of so-called economic policy uncertainty (EPU) indices in India to assess the degree of uncertainty faced by economic agents (Priyaranjan and Pratap (2020)). Such indices are basically constructed by setting up dictionaries that allow for the definition of specific terms that refer to uncertainty, and then searching them in the text considered (for instance in newspaper articles or on internet sites). These selected terms are then counted and aggregated to provide a synthetic index that reflects the degree of uncertainty displayed in the document of interest. Sentiment indices can be computed in this way, eg in order to measure the probability of the occurrence of financial instability episodes.

NLP is also helpful for policy evaluation. For instance, one can quantify the monetary policy stance that is communicated to the public via the publication of meeting minutes. Similarly, market expectations of interest rate decisions have been assessed by analysing market commentaries ahead of policy meetings in Indonesia (Andhika Zulen and Wibisono (2019)). Such exercises can be updated regularly, which is a key advantage compared to more traditional surveys of market participants. The information collected on economic agents' expectations can be particularly useful when future markets are not well developed, lack liquidity or are subject to unexpected shocks (Amstad and Tuazon(2020); Armas et al (2020)). By contrast, reported use of text data to inform financial stability policies has been relatively scarce so far, although it appears to be developing as well. Other applications using text analysis in Asian central banks have helped to: i) evaluate monetary policy credibility; ii) ensure consistency in central banks' communication of supervisory issues to financial institutions; iii) improve efficiency in the compilation of statistics (Chansang (2019)); iv) assess the state of the labor market (Bailliu et al (2019)) or of trade conditions (Amstad et al (2021)); v) extract information on tourism activities (popularity of travel destinations and potential associated topics); and vi) capture firms' sentiment or evaluate employees' feedback.[4]

Second, a large and increasing number of central banks support their economic analysis with **nowcasting models** drawing on big data. More than 40% of Asian central banks (24% in the rest of the world) indicated that big data is used for this purpose, especially to provide additional information on private consumption, industry/retail sales, retail/housing prices, payments and unemployment conditions (Graph 3, left panel). Matsumura et al (2021) combine GPS data with information on geographical coordinates of commercial and public facilities (such as shops and factories) to closely examine those sectors in which nowcasting can be applied to estimate (with a reported high level of precision and efficiency) household consumption and firm production. Finally, nowcasting models can help to fill statistical gaps, eg when reference series do not exist, are available only at a low frequency or are suddenly disrupted, as during the Covid-19 pandemic (De Beer and Tissot (2020)). This aspect has become particularly important for central banks, reflecting their dual role as producers as well as users of statistics.

For what specific purposes does your institution use big data? Graph 3



See footnote to Graph 1 for details on the sample composition.

Sources: IFC (2021a); authors' calculations.

Usually, these nowcasting exercises are frequently updated as new data come in, and various techniques – eg Lasso (Least Absolute Shrinkage and Selection Operator) – are applied to select the combination of variables that maximises the forecast at a given point in time (Richardson et al (2019)). One advantage is that this approach does not rely on specific relationships assumed ex ante (as is the case for bridge

4 Of course, economic agents adjust to new technologies. For example, Cao et al (2020) show that firms are aware that their filings are parsed and processed for sentiment via machine learning. Consequently, they avoid words that computational algorithms may perceive as negative. This could bias any analysis based on these filings.

models used for "traditional" nowcasting exercises) and may be better suited to identifying turning points, especially during times of economic upheaval (INSEE (2020)).

A third category includes the various applications developed by central banks to extract economy-wide insights from **granular financial data or other non-traditional sources of micro data.** Financial big data include large proprietary and structured data sets, such as those from trade repositories for derivatives transactions, or from credit registries for loans or individual payments. For instance, trade repositories' records have helped identify networks of exposures in Thailand (Chantharat et al (2017)). Similarly, information from credit registries have supported the assessment of credit quality, eg by improving estimates of default probabilities or loss-given-default (Pagano and Cappelli (1993)). And real-time gross settlement systems' data have helped help to show bank-firm interconnections through the payments processed.

In addition, special attention has been given to extract information from non-traditional data such as internet search queries like Google Trends that are supporting the monitoring exercises conducted by the Bank of Thailand (Sawaengsuksant (2019)). Other use cases of non-traditional sources include the analysis of: (i) electricity consumption to monitor the residential property market or export invoices to analyse the strength of the export sector in Malaysia (Wanitthanankun and Dummee (2017)); (ii) the number of job searches to monitor the evolution in the labor market in Thailand (Nuprae et al (2017)); (iii) mobile phone user traffic data to evaluate the effects of Covid-19 on mobility and migration (Chanthaphong and Tassanoonthornwong (2021)); (iv) patent applications by start-ups to estimate the economic impact of venture capital innovations in Japan (Washimi (2021)); and (vi) e-commerce sales (Yezekyan (2018)).

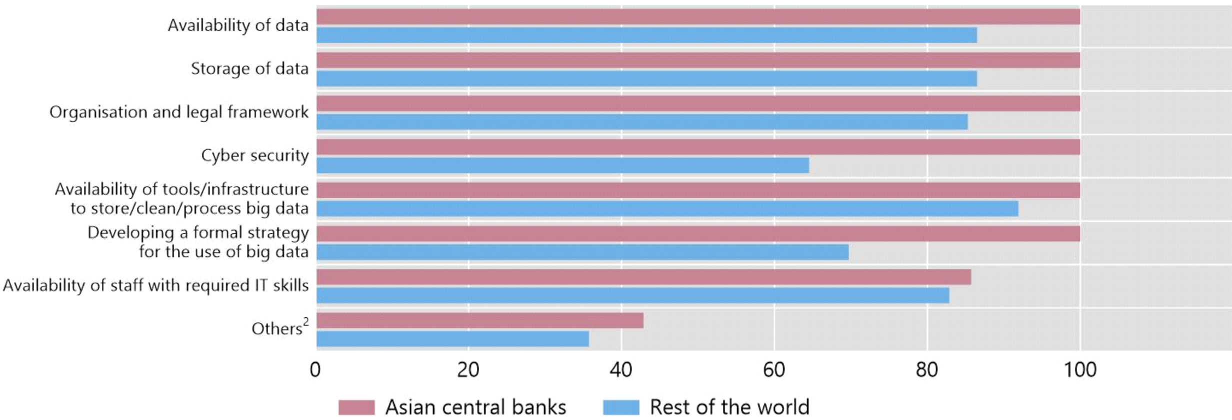
A fourth category comprises the **wide range of suptech and regtech applications** to support micro-supervisory tasks. This can cover multiple areas, as documented by Broeders and Prenio (2018), di Castri et al (2019), Coelho et al (2019) and Financial Stability Board (2020). In general, many of the applications developed among the Asian jurisdictions considered focus on micro-level risk assessment. For instance, firm-level information gathered from financial statements or newspapers can be used to support early warning exercises or enhance credit scoring (mentioned by about 55% and 45% of Asian central banks, respectively; Graph 3, right-hand panel). Another important area relates to fraud detection (almost 30% of the cases) – for instance, by screening credit contracts for suspicious terms and conditions to enhance consumer protection. Lastly, almost one third of surveyed Asian central banks deploy big data algorithms for anti-money laundering/combating the financing of terrorism (AML/CFT) purposes – for instance, when analysing payment transactions to identify suspicious patterns.

4. What are the main challenges in the use of big data?

As noted above, central banks and supervisory authorities in Asia already use extensively big data sources and analytics such as machine learning for research purposes, namely to inform monetary policy decisions, facilitate their statistical compilation work and support their regulatory and supervisory tasks. However, the use of big data poses **various challenges** for them. Graph 4 shows that these issues are actively discussed by central banks in Asia (in red), especially in comparison to their counterparts in the rest of the world (in blue). All the Asian central banks considered mention that they have active discussions on a wide range of topics, such as the availability of IT infrastructure, legal, security and privacy issues, as well as the availability and strategic use of big data. Interestingly, cyber security and the development of a formal strategy for the use of big data are areas that appear much more actively discussed compared to their counterparts in the rest of the world.

What is the focus of the discussions on big data within your institution? Graph 4

In per cent



[1] Respondents could select multiple options. See footnote to Graph 1 for details on the sample composition. [2] Includes "data quality and reliability", "data interpretation" and "data governance".

Sources: IFC (2021a); authors' calculations.

More specifically, the survey has highlighted five main challenges for Asian central banks in the use of big data. The first one is **setting up a reliable and high-powered IT infrastructure** (IFC (2020)). Providing adequate computing power and software involves high up-front costs. Many central banks have undertaken important initiatives to develop big data platforms to facilitate the storage and processing of large and complex data sets. One possible approach is represented by so-called data lakes, obtained from pooling different data sets that are curated for future use. A reliable and safe IT infrastructure is a prerequisite not only for big data analysis, but also to prevent cyberattacks.

Second, central banks need to **build up human capital** to exploit big data. Setting up and maintaining big data platforms requires a specific type of skillset, combining statistical, IT, and analytical/mathematical aspects. Yet the supply of "data scientist" is scarce and they are in high demand (Cœuré (2020)), in both the public and the private sector. One solution is for central banks to train existing staff but learning the new techniques that are needed can require significant time and effort. In addition, experience shows that these skill adjustments should take place beyond the operational level, eg the statisticians in charge of using advanced tools. Those analysing the output of complex models must also have a good understanding of the new techniques in order to ensure that big data predictions are not only accurate but also representative and "interpretable" – so that specific explanatory causes or factors can be identified and communicated for policy use. Another issue is attracting and retaining talent, especially in the face of intense competition from the private sector, as well as from advanced economies especially for the less developed jurisdictions in Asia. This may also call for a review of existing public compensation schemes, career systems and internal hierarchical organisations in central banks.

A third challenge are the **legal underpinning and ethical aspects** for the use of private and confidential data. Reputational aspects may hinder the use of information sourced from the internet when little is known about its accuracy and the respect of methodological standards that central banks have to comply with, not least in view of the key role they play in National Statistical Systems. For instance, internet-based indicators such as search queries and messages on social media may not be representative of the real economy – not everybody is on Twitter, or only a subset of the CPI basket prices can be scraped from the web. Moreover, various terms and conditions may restrict the use of these data and certain forms of web-scraping are illegal in some jurisdictions. In general, web crawlers cannot obtain data from sites that require authentication.

Considering ethics and privacy aspects, citizens might feel uncomfortable with the idea that central banks are scrutinising their search histories, social media postings or listings on market platforms. While these concerns are not new, the amount of data produced in a mostly unregulated environment makes them more urgent (Jones and Tonetti (2020), Boissay et al (2020)). Certainly, when US consumers were asked in a systematic survey whom they trust with safeguarding their personal data, the respondents reported that they trust big techs the least (Armantier et al (2021)). They had in fact far more trust in traditional financial institutions, followed by government agencies and fintechs. Similar patterns are present in Asian countries (Chen et al (2021)).[5] Yet, ensuring privacy against unjustified intrusion not only by commercial actors but also by government has the attributes of a basic right. For these reasons, the issue of data governance has emerged as a key public policy concern (IFC (2021b)).

A fourth challenge is "**algorithmic fairness**". This consideration can be less relevant for some tasks (eg nowcasting), but it may matter greatly for others (eg evaluating the suitability of regtech applications), and in general any application of machine learning that effects individuals would need to be subject to fairness validations (MacCarthy (2019)). A main issue is that algorithms are often trained on pre-classified data sets that can be subject to (known or unknown) biases, including related to gender and ethnicity.[6] Moreover, the relationship that seems to exist between unstructured data and a certain phenomenon may unexpectedly deteriorate when additional information arrives (eg the incorporation of new, "out-of-sample" information).The failure of Google Flu Trends provides a good example of these perils, as it was initially intended to provide estimates of influenza activity based on Google Search queries but was discontinued in the mid-2010s (Lazer et al (2014)).

Finally, **data quality** issues are also significant, since much of the new big data collected as a by-product of economic or social activities needs to be curated before

5 IIF (2020) finds that there is no "one-size-fits-all" approach to machine learning governance, and there are interesting regional differences, many of which can be attributable to existing non-discrimination and data protection laws.

6 For instance, data on past loan applications could reflect any discriminatory decisions on the part of loan officers vis-à-vis minorities or women (Angwin et al (2016), Ward-Foxton (2019)). Likewise, unrepresentative data could lead an algorithm to wrongly infer attributes about underrepresented segments of the population or perpetuate any previous biases.

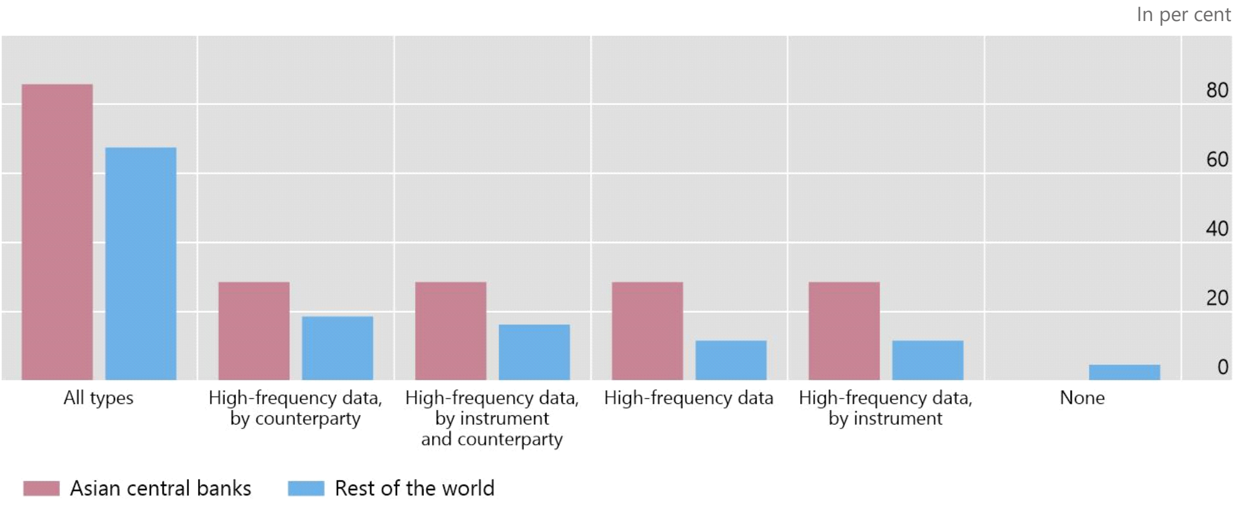
proper statistical analysis can be conducted. This stands in contrast to traditional sources of official statistics that are designed for a specific purpose, eg surveys and censuses. Major challenges include data cleaning (eg for sources like newspapers, social media or financial big data records), sampling and representativeness (eg in the case of Google searches or employment websites) and matching new data to existing sources, as documented by Siksamat (2021) in the case of Thailand.

5. Is there a role for policy cooperation?

Cooperation could foster central banks' use of big data, in particular through collecting and showcasing successful projects and facilitating the sharing of experiences. For instance, developing technical discussions between institutions is seen as a good way to build the necessary skillset among staff and develop relevant IT tools and algorithms that are best suited to central banks' needs.

Looking ahead, a promising area for collaboration among central banks in Asia could be in **global payments data**. More than 85% of Asian central banks reported an active use of high frequency payment data in their institutions, with a primary focus on either the type of instruments, counterparties involved or both. This ratio is much higher compared with other central banks in the rest of the world (about 65%; Graph 5). Moreover, all of Asian central banks expressed interest in contributing to a pilot study on payment data (Graph 6, left-hand panel), especially to develop surveillance exercises with a focus on interconnectedness in the financial system. This stands in contrast to their counterparts in the rest of the world, where interest in using payment data is primarily limited to nowcasting purposes (right-hand panel).

Which types of payments data are useful for your institution? Graph 5

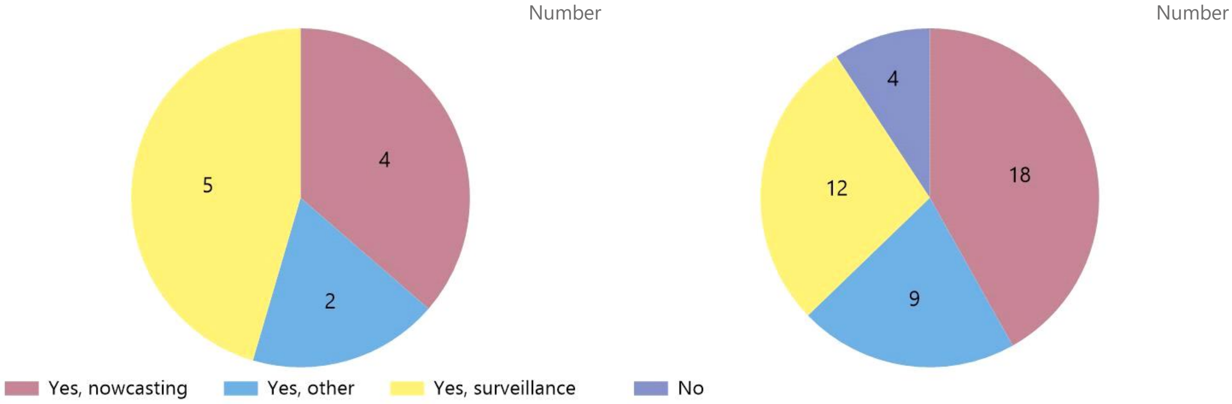


See footnote to Graph 1 for details on the sample composition.

Sources: IFC (2021a); authors' calculations.

Would your institution be willing to contribute to a pilot study on the use of payments data? Graph 6

Asian central banks | Rest of the world



See footnote to Graph 1 for details on the sample composition (note: only 29 IFC members responded to this specific question); multiple answers possible in each jurisdiction.

Sources: IFC (2021a); authors' calculations.

International financial institutions can foster cooperation around big data. For instance, they can help develop in-house big data knowledge, helping to reduce central banks' reliance on big data services providers, which can be expensive and entail significant legal and operational risks. Indeed, the IFC has been actively supporting such exchange of experience at the global level, and several complementary initiatives are being developed in the Asian region, for instance among EMEAP central banks.[7]

International bodies can also facilitate innovation by **promoting technological solutions and initiatives** to enhance the global statistical infrastructure. In this regard, the BIS Innovation Hub has identified as strategic priorities, among others, effective supervision (including regtech/suptech) and open banking/finance that could benefit from drawing on big data sources and tools. It is currently developing its work program in these fields, with a view to producing proofs of concept (PoC) that can benefit the central banking community.

Initial projects in the field of particular relevance for Asian central banks include Ellipse, led by the Singapore Centre of the Hub, and Genesis (Hong Kong Centre). Ellipse is a PoC that aims to demonstrate the functionalities and feasibility of an integrated regulatory data and analytics platform that can (i) reduce compliance burdens placed on financial institutions by moving away from template-based regulatory reporting requests; (ii) be nearer to "real-time" and relevant to current events to support supervisory judgments and actions, both locally and globally; (iii) support a move towards newer digitally enabled architectures to replace traditional concepts and processes of data collection; and (iv) enable predictive insights and early warning by integrating big data analytics. Turning to Genesis, this project explores

[7] Executives' Meeting of East Asia-Pacific Central Banks (EMEAP) is a co-operative forum of eleven central banks and monetary authorities in the East Asia and Pacific region. Common projects are developed in the areas of banking supervision and resolution, financial markets, payments and market infrastructure, and information technology.

the "green art of the possible" through combining blockchain, smart contracts, digital assets, and the internet-of-things. The underlying vision is that an investor can download an app to invest into government bonds, so that the proceeds can be used to develop a green project. Over the bond's lifetime, the investor would be able to not just see accrued interest, but also track in real time how much clean energy is being generated, and the consequent reduction in CO2 emissions linked to the individual investment.

6. Conclusion

The world is changing and so is the way it is measured. This paper provides an overview of the use of big data in the Asian central bank community. It leverages on a survey conducted in 2020 among the members of the IFC. The specific responses from seven Asian central banks were analysed and compared with those of other central banks in the rest of the world. The overall picture suggests that, while central banks in other regions see similar challenges and opportunities in the use of big data, those located in Asia have very distinctive features.

First, Asian central banks define big data in an encompassing way that includes not only unstructured, non-traditional data but also structured data sets to a larger extent compared to other regions. Second, interest in big data appears higher in Asia, including at the senior policy level. Third, a large majority of Asian central banks report dealing with big data to support economic research, monetary and financial stability policies as well as their statistical production tasks, a ratio that is slightly above the situation reported in other regions. The related big data projects are developed mainly in the areas of NLP, nowcasting, applications to extract economy wide insight, and suptech/regtech solutions. Fourth, the advent of big data poses new challenges, such as the reliability of IT infrastructures, legal aspects around privacy, algorithmic fairness, and data quality. Interestingly, there is a somewhat higher interest among Asian central banks for analysing these issues, with topics such as cyber security and the development of a formal strategy for the use of big data being particularly high on their agendas.

Asian (and other) central banks are willing to join forces to reap the benefits of big data, the IFC survey shows. International financial institutions can support these cooperative approaches.[8] They can facilitate innovation by promoting technological solutions to harmonise data standards and processes among jurisdictions, and important projects have been already launched in Asia.

8 Specific initiatives to foster closer collaboration and accelerate innovation efforts include the ASEAN Open Data Dictionary, ASEANstats, Asia Open data Partnership (Dataportal.Asia).

References

Amstad, M, G Cornelli, L Gambacorta and D Xia (2020): "Investors' risk attitude in the pandemic and the stock market: new evidence based on internet searches", *BIS Bulletin,* no 25.

Amstad, M, L Gambacorta, C He and D Xia (2021): "Trade sentiment and the stock market: new evidence based on big data textual analysis of Chinese media", *BIS Working Papers,* no 917.

Andhika Zulen, A and O Wibisono (2019): "Measuring stakeholders' expectation on central bank's policy rate", *IFC Bulletin,* no 49.

Angwin, J, J Larson, S Mattu and L Kirchner (2016): "Machine bias", *ProPublica*.

Armantier, O, S Doerr, J Frost, A Fuster and K Shue (2021): "Whom do consumers trust with their data? US survey evidence", *BIS Bulletin,* no 42.

Armas, J C A and P K A Tuazon (2020): "Revealing investors' sentiment amid COVID-19: the big data evidence based on internet searches", *BSP Working Paper Series*, July.

Bailliu J, X Han, M Kruger, Y Liu and S Thanabalasingam (2019): "Can media and text analytics provide insights into labor market conditions in China?", *IFC Bulletin,* no 49.

Bank of Japan (2020): "Impact of COVID-19 on private consumption", *Outlook for Economic Activity and Prices*, July, Box 3.

Boissay, F, T Ehlers, L Gambacorta and H S Shin (2020): "Big techs in finance: on the new nexus between data privacy and competition", in: Rau R, Wardrop R and Zingales L (eds), *The Handbook of Technological Finance*, Palgrave Macmillan.

Broeders, D and J Prenio (2018): "Innovative technology in financial supervision (suptech) – the experience of early users", *FSI Insights,* no 9.

Buch, C (2019): "Building pathways for policy making with big data", welcoming remarks at the International seminar on big data, *IFC Bulletin,* no 50.

Cao, S, W Jiang, B Yang and A Zhang (2020): "How to talk when a machine is listening: corporate disclosure in the age of AI", *NBER Working Paper,* no 27950.

Chansang, P (2019): "Data management in the data evolution era at Bank of Thailand", *IFC Bulletin,* no 53.

Chanthaphong, S and T Tassanoonthornwong (2021): "Workers' mobility and Covid 19 pandemic: An analysis using mobile big data", *Bank of Thailand articles*.

Chantharat, S, A Lamsam, K Samphantharak and P Tangsawadirat (2017): "A new perspective on Thai household debt through credit bureaus' big data", *PIER discussion paper,* October.

Chen, S, S Doerr, J Frost, L Gambacorta and H S Shin (2021): "The fintech gender gap", *BIS Working Papers,* no 931.

Coelho, R, M De Simoni and J Prenio (2019): "Suptech applications for anti-money laundering", *FSI Insights,* no 18.

Cœuré, B (2017): "Policy analysis with big data", speech at the conference on Economic and Financial Regulation in the Era of Big Data, organised by the Bank of France, Paris.

——— (2020): "Leveraging technology to support supervision: challenges and collaborative solutions", speech at the *Financial Statement event series*, Peterson Institute for International Finance.

Cornelli, G, J Frost, L Gambacorta, R Rau, R Wardrop and T Ziegler, (2020): "Fintech and big tech credit: a new database", *BIS Working Papers,* no 887.

De Beer, B and B Tissot (2020): "Implications of Covid-19 for official statistics: a central banking perspective", *IFC Working Papers*, no 20.

di Castri, S, S Hohl, A Kulenkampff and J Prenio (2019): "The suptech generations", *FSI Insights*, no 19.

Doerr, S and L Gambacorta (2020a): "Identifying regions at risk with Google Trends: the impact of Covid-19 on US labor markets", *BIS Bulletin,* no 8.

——— (2020b): "Covid-19 and regional employment in Europe", *BIS Bulletin*, no 16.

Doerr, S, L Gambacorta and J M Serena (2021): "Big data and machine learning in central banking", *BIS Working Papers*, no 930.

Draghi, M (2018): Welcome remarks at the third annual conference of the ESRB.

Financial Stability Board (2020): "The use of supervisory and regulatory technology by authorities and regulated institutions. Market developments and financial stability implications", Report to the G20.

Forbes (2012): "The Age of Big Data", accessed 12 June 2020.

Institut national de la statistique et des études économiques (INSEE) (2020): " "High-frequency" data are especially useful for economic forecasting in periods of devastating crisis", *Point de Conjoncture*, June, pp 29–34.

Institute of International Finance (2020): "Machine learning governance".

Irving Fisher Committee (2017): "Big data", *IFC Bulletin*, no 44.

——— (2020): "Computing platforms for big data analytics and artificial intelligence", *IFC Report*, no 11.

——— (2021a): "Use of big data sources and applications at central banks", *IFC Report*, no 13.

——— (2021b): "Issues in Data Governance", *IFC Bulletin*, no 54.

Jones, C and C Tonetti (2020): "Nonrivalry and the economics of data", *American Economic Review*, 110(9), pp 2819–58.

Lazer, D, R Kennedy, G King and A Vespignani (2014): "The parable of Google Flu: traps in big data analysis", *Science*, 343(6176), pp 1203–05.

MacCarthy, M (2019): "Fairness in algorithmic decision-making", Brookings Institution's Artificial Intelligence and Emerging Technology (AIET) Initiative, Brookings Institution.

Matsumura, K, Y Oh, T Sugo and K Takahashi (2021): "Nowcasting economic activity with mobility data", *Bank of Japan Working Paper Series*, no 21-E-2.

Mehrhoff, J (2019): "Demystifying big data in official statistics – it's not rocket science!", *IFC Bulletin*, no 49.

Nuprae, W, W Nakwatara and P Sawangsuk (2017): "What can big data tell about the Thai labor market?", *PIER discussion paper*, no 9, Puey Ungphakorn Institute for Economic Research.

Nymand-Andersen, P (2016): "Big data: the hunt for timely insights and decision certainty", *IFC Working Papers*, no 14.

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

Pagano, M and T Jappelli (1993): "Information sharing in credit markets", *Journal of Finance*, no 48(5), pp 1693–18.

Priyaranjan, N and B Pratap (2020): "Macroeconomic effects of uncertainty: a big data analysis for India", *RBI Working Paper*, no 4.

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

Sawaengsuksant, P (2019): "Standardised approach in developing economic indicators using internet searching applications", *IFC Bulletin*, no 50.

Siksamat, S (2021): "Collecting data: new information sources", *IFC Bulletin*, no 54.

Tissot, B (2017): "Big data and central banking", *IFC Bulletin*, no 44.

——— (2019): "Financial big data and policy work: opportunities and challenges", *Eurostat Statistical Working Papers*, no KS-TC-19-001-EN-N.

Wanitthananankun, J and J Dummee (2017): "Micro data usage enhancement in Bank of Thailand", *MyStats 2017*, Department of Statistics Malaysia.

Ward-Foxton, S (2019): "Reducing bias in AI models for credit and loan decisions".

Washimi, K (2021): "Venture capital and startup innovation – big data analysis of patent data", *Bank of Japan reports and research papers*, Bank of Japan.

Yezekyan, L (2018): "Compilation of e-commerce data for balance of payments statistics", *IFC Bulletin*, no 48.

Appendix: Big data projects in Asian central banks

Central Bank	Project	Data source	Purpose	Platform
Bangko Sentral ng Pilipinas	BSP Big Data Project	Government/ Academia	Develop the big data Roadmap and big data governance framework; operationalise big data system Prototypes	R, Python, Geoda, QGIS
	Enterprise Data Warehouse	Data Warehouse solution/service-providers	Have a single database for the BSP	
	Anomaly Detection in Data	Reports from BSP supervised and unsupervised entities	For financial stability, anti-money laundering, and fraud detection purposes	SAS, Python
Bank Indonesia	Indicator of job demand from online job vacancy portals	Online job vacancy portals	Produce proxy indicator/nowcasting employment	Hadoop, Hive, Spark, Impala
	Identification of main counterparties in forex market	RTGS	Identify main counterparties in forex market from payment system data	Hadoop, Hive, Spark, Impala
	Indicator of consumption from payment system data	Clearing system	Produce indicator of consumption (household and government) from payment system data	Hadoop, Hive, Spark, Impala
	Indicator of property prices from online property portals	Online property portals	Produce statistics for property prices in secondary market	Hadoop, Hive, Spark, Impala
	Indicator of automobile supply from online automobile portals	Online automobile portals	Produce proxy indicator/nowcasting automobile supply	Hadoop, Hive, Spark, Impala
	Analysis of travelers' reviews from online travel portals	Online travel portals	Produce analysis of popularity of travel destinations and their main issues	Python
	Indicator of e-commerce sales	E-commerce sites	Produce proxy indicator of household consumption, retail sales, and use of payment instruments	Hadoop, Hive, Spark, Impala
	Indicator of Economic Policy Uncertainty	News articles	Produce indicator of Economic Policy Uncertainty for Indonesia	Python
	Indicator of monetary policy credibility	News articles	Produce indicator of public's perception of monetary policy credibility	Python
	Interconnectedness of banks in payment system	RTGS	Identify core and periphery banks in payment system	Hadoop, Hive, Spark, Impala
	Interconnectedness of corporations in payment system	RTGS	Identify network structure of corporations	Hadoop, Hive, Spark, Impala
Bank of Japan	A Network Analysis of the Repo JGB Market	Collected from financial institutions located in Japan	Analyse the structure of the Japanese repo market	R
	Release of "Statistics on Securities Financing Transactions in Japan"	Collected from financial institutions located in Japan	Publish data on securities financing transactions in Japan	
	Corporate behavior and innovation	Japan's patent data provided by Panasonic system solutions	Analyse the effects of R&D investment on productivity growth	R
	Analysis of business and consumer sentiments	Economy Watchers Survey	Analyse business and consumer sentiments using comments from respondents of the survey using text analysis	R, Python

Bank of Thailand	Leading indicators for export	Thai Customs Department	Develop leading indicators	Python
	Manufacture sector structure	Manufacturing firm census from NSO	Understand the structure of manufacturing sectors	Stata
	High-rise residential property occupancy rate	Electricity bills, Provincial Electricity Authority	Monitor real demand for high-rise property	RStudio
	Use internet search technology	Google Trends/Correlate	Develop indicators to help monitor economic conditions	Google Trends/Correlate
	Use text analytics to improve operation	Comptroller General's Department	Use text analytics to improve efficiency of statistics compilation	Python
	SME financing behavior and SME credit risks	Credit registry data/ micro data obtained for supervisory purposes	Identify SMEs' viability and assess credit risks	Impala/ Tableau
	Export indicator from data analytics	Thai Customs Department and Bank of Thailand	Develop indicator for monitor Thai exports	Python
	Stylised facts on invoicing currency and natural hedge of Thai exporters	Thai Customs Department and Bank of Thailand	Explore invoice structure and natural hedge of Thai exporters	RStudio
	Self-employed labor income	Labor Force Survey from NSO	Determine self-employed labor income to monitor economy	RStudio and Stata
	Job switching pattern of labor	Social Security Office	Explore and understand the job switching behavior	RStudio and Stata
	Structure of retail trades	Web-scraping, firm balance sheet, Labor Force Survey	Understand the structure of retail trade sectors	Stata and Tableau
Bank Negara Malaysia	Credit modelling for retail and non-retail borrowers	Internal credit registry database	Predict probability of default, loss given default etc.	R programming
	News monitoring and sentiment analysis dashboard	Public news sites	Enhance surveillance of topics of interest and understand public sentiment on these topics	Python, Django, ElasticSearch etc.
	Analytical solution for analysis of AML/CFT-related data	Data submitted by regulated entities, internal databases	Construct network models to establish relationships between entities and provide search capability	Python, Django, ElasticSearch, Neo4j etc.
	Employee feedback text analysis	Internal talent management surveys	Analyse employees' key feedback from talent management surveys	Python, Django, HuggingFace
	Supervisory letter text analysis	Internal data	Ensure consistency in the communication of supervisory issues to financial institutions	Python, Django, HuggingFace, ElasticSearch
Reserve Bank of India	Centralised Information Management System	Structured data from regulated entities and unstructured web-scraped data	Create single repository comprising structured and unstructured big data, and use it for analyses	End-to-end Hadoop eco-system, integrated with R / Python
	Outlook on specific economic indicators based on media articles	Online Portals	Big data analytics, ML and related techniques	Hadoop / R / Python
	Food inflation based on online retail prices	Online Portals	Big data analytics, ML and related techniques	Hadoop / R / Python
	Housing Price Index based on online property advertisements	Online Portals	Big data analytics, ML and related techniques	Hadoop / R / Python

Source: IFC (2021a).