JCER Working Paper
AEPR series
No. 2021-2-4

This is the pre-peer- reviewed version of the following article: "Big Data in Asian Central Banks", *Asian Economic Policy Review*, vol. 17, issue 2, which has been published in final form at https://onlinelibrary.wiley.com/doi/10.1111/aepr.12376 and DOI: 10.1111/aepr.12376.

Big Data in Asian Central Banks

Giulio Cornelli
Sebastian Doerr
Leonardo Gambacorta and
Bruno Tissot
(Bank for International Settlements)

This paper was prepared for the Twenty-fourth Asian Economic Policy Review (AEPR) Conference "Digital Finance" on October 1, 2021, via zoom.

July 2022

Asian Economic Policy Review
Japan Center for Economic Research

To authors

If you want to introduce the same working paper you wrote and presented at the AEPR conference held via zoom on Oct. 1, 2021, in your own/your affiliation's website, please be aware the following requirements.

To ensure that all citations and references to your published article are captured by the SSCI (Social Sciences Citation Index), authors are required to amend the cover page of your working paper as soon as practical after publication in AEPR. The amended cover page should include the full article citation, journal name, volume and issue, and DOI, as well as a hyperlink to the published article. The cover page of JCER Working Paper AEPR series has been already amended after publication in AEPR. The face of this working paper is an example of an amended working paper cover page.

Big data in Asian central banks

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

Bank for International Settlements

Research highlights

1. Asian central banks define big data in an encompassing way that includes unstructured non-traditional as well as structured data sets.
2. Interest in big data is particularly high in Asia, with projects developed in NLP, nowcasting, applications to extract economy insights, and suptech/regtech applications.
3. The advent of big data poses new important challenges, with specific attention paid by Asian central banks to cyber security and data strategy.
4. Cooperation among public authorities in Asia is seen as instrumental also to facilitate the use of payments data and promote innovative technological solutions.

Abstract

This paper reviews the use of big data in Asian central banks, leveraging on a survey conducted among the members of the Irving Fischer Committee. The analysis reveals four main insights. First, Asian central banks define big data in a more encompassing way. Second, interest in big data appears higher in Asia, including at the senior policy level. Third, Asian central banks report dealing with big data to support a wide range of tasks. Fourth, big data poses new challenges and increases the need for international policy cooperation, especially to make 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

* Corresponding author: Leonardo Gambacorta, Bank for International Settlements, Centralbahnplatz, 2 4002, Basel Switzerland (email: leonardo.gambacorta@bis.org).

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 digitization, with the development of the "internet of things" and the 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 2018, 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 reflects 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.

Looking ahead, 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. To this end, we use the responses from seven Asian central banks and compare them with those of other central banks in the rest of the world.[1]

The survey focused on 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 central banks facing in the use of big data and how can they be overcome? Based on answers to the survey, this paper contrasts Asian central banks with their peers in the rest of the world.

The results of the survey for Asian central banks uncover 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 "non-traditional" (or unstructured) data often characterized 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" (i.e. structured) data sets. These can be the result of explicit reporting requirements set by public regulators; they are also often "organic" by-products collected as a result of commercial (e.g. payment transactions), financial (e.g. tick-by-tick price quotes observed in financial markets) and administrative (e.g. 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 reported 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 countries.

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 slightly above 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 challenge is setting up a reliable and high-powered IT infrastructure. While many Asian central banks 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. 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 (e.g., data scientists). Other challenges include the legal basis for using private data and the safety, ethics 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 the rest of the world.

The rest of the paper is organized as follow. Section 2 provides an overview of how Asian central banks define big data. Sections 3 illustrates in which fields Asian central banks 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 analyzing 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".[2]

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 Figure 1, only 15% of the respondents define big data exclusively as large non-traditional or unstructured data that require new techniques for the analysis (in contrast, 30% of their counterparts in the rest of the world have such a narrow definition). The remaining 85% of Asian respondents also include traditional and structured data sets in their definition of big data. These structured data sets include 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 analyzed. The reason for this is, in part, that traditional statistical techniques face hurdles when applied to unstructured data. For instance, to analyze handwritten text, it must first be turned into structured data, as is done for instance with NLP algorithms.

[Figure 1 here]

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 – “the internet of things” – may not necessarily be “big”, but it is complex and cannot be easily analyzed 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 organized 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). 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 analyzed, they must be cleaned and curated, i.e. organized 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. Among Asian central banks, the share has risen from 33% to 86%. Looking specifically into Asia, around 60% of central banks reported that they discuss big data issues extensively, a ratio that is significantly the one (42%) observed in the rest of the world. All of these Asian respondents indicated a high to very-high level of interest at the senior policy level, compared only 58% in the rest of the world.

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 Figure 2) appear to use big data in most areas by more than their peers (blue bars), except for research purposes. In particular, they make greater use of non-traditional data (darker bars) to support financial stability and monetary policy. They also use it for supervision and regulation (suptech and regtech), often reflecting their specific mandates.

[Figure 2 here]

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 annex.

A first type of application uses textual information through NLP. The goal is generally to turn qualitative text-based information 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 & 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 are also computed 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 analyzing market commentaries ahead of policy meetings in Indonesia (Anthia Zulen & 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 market expectations can be particularly useful when future markets are not well developed, lack liquidity or are subject to unexpected shocks (Amstad et al, 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 help 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, 2020) or of trade conditions (Amstad et al, 2021); v) extract information for tourism (popularity of travel destinations and potential issues); vi) extract firms' sentiment or evaluate employees' feedback.[3]

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, house prices and unemployment conditions (Figure 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 high level of precision and efficiency) household consumption and firm production. Finally, nowcasting models can help to fill statistical gaps, e.g. when reference series do not exist or are available only at a low frequency or are suddenly disrupted, as during the Covid-19 pandemic (De Beer & Tissot 2020). This aspect as become particularly important, reflecting the dual role played by central banks as producers as well as users of statistics.

[Figure 3 here]

Usually, these nowcasting exercises are frequently updated as new data come in, and various techniques – e.g. Lasso (Least Absolute Shrinkage and Selection Operator) – are applied to select the combination of variables that maximizes 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 of bridge 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 form **granular financial data or other non-traditional sources for 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 have helped identify networks of exposures in Thailand (Chantharat et al. 2017). Credit registries support the assessment of credit quality, e.g. by improving estimates of default probabilities or loss-given-default (Pagano & Jappelli 1993). Real-time gross settlement system data help to show bank-firm interconnections through their payments. A special attention is 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 sources of non-traditional data include the analysis of: (i) electricity consumption to monitor the residential property market or export invoices to analyze the strength of the export sector in Malaysia (Wanitthanankun & 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 to evaluate the effects of Covid-19 on mobility and migration (Chantapong &

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 policies. This can cover multiple tasks, as documented by Broeders and Prenio (2018), di Castri et al. (2019), Coelho et al. (2019) and Financial Stability Board (2020). In general, among the Asian jurisdictions considered, many of these applications 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 55% and 45% of Asian central banks, respectively; Figure 3, right-hand panel). Another important area relates to fraud detection (around 30% of the cases) – for instance, by screening credit contracts for suspicious terms and conditions to enhance consumer protection. Lastly, almost one third of Asian central banks deploy big data algorithms for anti-money laundering/combating the financing of terrorism (AML/CFT) purposes – for instance, when analyzing 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, to inform monetary policy decisions, to facilitate their statistical compilation tasks and for regulation and supervision. However, the use of big data poses various challenges for them. Figure 4 shows that these topics are actively discussed by Asian central (in red) banks, 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 and human capital, 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 topics that appear much more actively discussed compared to their counterparts in the rest of the world

[Figure 4 here]

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**. Providing adequate computing power and software, as well as training existing or hiring new staff, 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, e.g. the statisticians in charge of using advanced tools; those analyzing the output of complex models must also have a good understanding of 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 organizations 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. 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 scrutinizing 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 & Tonetti 2020, Boissay et al. 2020). Beyond the economic consequences, ensuring privacy against unjustified intrusion by both commercial and government actors has the attributes of a basic right. For these reasons, the issue of data governance has emerged as a key public policy concern. 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 have far more trust in traditional financial institutions, followed by government agencies and fintechs. Similar patterns are present in Asian countries (Chen et al. 2021).[4]

A fourth challenge is "**algorithmic fairness**". Considerations of algorithmic fairness are less relevant for some tasks (e.g. nowcasting), but they may matter greatly for others (e.g. evaluating the suitability of regtech applications), and in general any application of machine learning that effects individuals' needs to be subject to fairness validations (MacCarthy 2019). Algorithms train on pre-classified data sets that can be subject to biases, including related to gender and ethnicity.[5] Moreover, the relationship that seems to exist between unstructured data and a certain phenomenon may unexpectedly deteriorate when additional information arrives (e.g. 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 proper statistical analysis can be conducted. This stands in contrast to traditional sources of official statistics that are designed for a specific purpose, e.g. surveys and censuses. Major challenges include data cleaning (e.g. in the case of media, social media or financial data), sampling and representativeness (e.g. 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. In particular, developing technical discussions between institutions is seen as a promising way to build the necessary skillset among staff and develop relevant IT tools and algorithms that are best suited to central banks' needs.

[Figure 6 here]

Looking ahead, a promising area for collaboration among central banks in Asia could be in global payments data. Around 90% of Asian central banks reported an active use of high frequency payment data in their institutions, with a primary focus on the type of instruments and counterparties involved. This ratio is much higher compared with other central banks in the rest of the world (60%; Figure 5). Moreover, all of them expressed interest in contributing to a pilot study on payment data (Figure 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 mostly limited to nowcasting purposes (right-hand panel).

International financial institutions can foster cooperation around big data. For instance, they can help develop in-house big data knowledge, reducing central banks' reliance on big data services providers, which can be expensive and entail significant legal and operational risks. Indeed, the IFC Committee has been actively supporting such exchange of experience at the global level, and several complementary initiatives are being developed in Asian region for instance among EMEAP central banks.[6]

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 data platforms/open finance that could draw on big data. 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. Turing to Genesis, this project explores 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 CO_2 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 Irving Fischer Committee. We use the specific responses from seven Asian central banks and compare them with those of other central banks in the rest of the world. The overall picture suggests that, while central banks all over the world see similar challenges and opportunities in the use of big data, those located in Asian region 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 applications. Fourth, the advent of big data poses new challenges, such as 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 analyzing these issues, and cyber security and the development of a formal strategy for the use of big data are topics that are 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.[7] They can facilitate innovation by promoting technological solutions to harmonize data standards and processes among jurisdictions, and important projects have been already launched in Asia.

Acknowledgments

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 an Helio Vale. The views expressed are those of the authors and not necessarily those of the BIS.

References

Amstad M., Cornelli G., Gambacorta L. & Xia D. (2020). Investors' risk attitude in the pandemic and the stock market: new evidence based on internet searches. BIS Bulletin 25.

Amstad M., Gambacorta L., He C. & Xia D. (2021). Trade sentiment and the stock market: new evidence based on big data textual analysis of Chinese media. BIS Working Papers 917.

Andhika Zulen A. & Wibisono O. (2019). Measuring stakeholders' expectation on central bank's policy rate. IFC Bulletin 49, Irving Fisher Committee on Central Bank Statistics.

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

Armantier O., Doerr S., Frost J., Fuster A. & Shue K. (2021). Whom do consumers trust with their data? US survey evidence, BIS Bulletins 42.

Armas J. C. A. & Tuazon P. K. A. (2020). Revealing investors' sentiment amid COVID-19: the Big Data evidence based on internet searches, BSP Working Paper Series 2020-07.

Bailliu J., Han X., Kruger M., Liu Y. & Thanabalasingam S. (2019). Can media and text analytics provide insights into labor market conditions in China?. IFC Bulletin 49.

Bank of Japan (2020). Impact of COVID-19 on Private Consumption. Outlook for Economic Activity and Prices, July, Box 3.

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

Broeders D. & Prenio J. (2018). Innovative technology in financial supervision (suptech) – the experience of early users. FSI Insights 9.

Buch C. (2019). Welcoming remarks International seminar on big data, building pathways for policy-making with big data. IFC Bulletin 50.

Cao S., Jiang W., Yang B. & Zhang A. (2020). How to talk when a machine is listening: corporate disclosure in the age of AI. NBER Working Paper 27950.

Chansang P. (2019). Data management in the data evolution era at Bank of Thailand. IFC Bulletin 53.

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

Chantharat S., Lamsam A., Samphantharak K. & Tangsawadirat P. (2017). A new perspective on Thai household debt through credit bureaus' Big Data. PIER discussion paper 10/2017.

Chen, S., Doerr S., Frost J., Gambacorta L. & Shin H. S. (2021). The fintech gender gap. BIS Working Papers 931.

Coelho R., De Simoni M. & Prenio J. (2019). Suptech applications for anti-money laundering. FSI Insights 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.

Cœuré B. (2020). Leveraging technology to support supervision: challenges and collaborative solutions. Speech at the Financial Statement event series, Peterson Institute for International Finance.

Cornelli, G., Frost, J., Gambacorta, L., Rau, R., Wardrop, R. & Ziegler, T. (2020). Fintech and big tech credit: a new database. BIS Working Paper 887.De Beer B. & Tissot B. (2020). Implications of Covid-19 for official statistics: a central banking perspective. IFC Working Paper 20.

di Castri S., Hohl S., Kulenkampff A. & Prenio J. (2019). The suptech generations. FSI Insights 19.

Doerr S. & Gambacorta L. (2020a). Identifying regions at risk with Google Trends: the impact of Covid-19 on US labor markets. BIS Bulletin 8.

Doerr S. & Gambacorta L. (2020b). Covid-19 and regional employment in Europe. BIS Bulletin 16.

Doerr S., Gambacorta L. & Serena J. M. (2021). Big data and machine learning in central banking. BIS Working Paper 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 2020, 29–34.

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

Irving Fisher Committee (2017). Big data. IFC Bulletin 44.

Irving Fisher Committee (2020). Computing platforms for big data analytics and artificial intelligence. IFC Report 11.

Irving Fisher Committee (2021a). Use of big data sources and applications at central banks. IFC Report 13.

Irving Fisher Committee (2021b). Issues in Data Governance. IFC Bulletin no. 54.

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

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

Matsumura K., Oh Y., Sugo T. & Takahashi K. (2021). Nowcasting Economic Activity with Mobility Data. Bank of Japan Working Paper 21-E-2, Bank of Japan Working Paper Series.

Mehrhoff J. (2019). Demystifying big data in official statistics – it's not rocket science!. IFC Bulletin 49.

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

Nuprae W., Nakwatara W. & Sawangsuk P. (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. & Pantelidis E. (2018). Google econometrics: nowcasting euro area car sales and big data quality requirements. European Central Bank Statistics Paper 30.

Pagano M. & Jappelli T. (1993). Information sharing in credit markets. Journal of Finance, 48(5), 1693–18.

Priyaranjan N. & Pratap B. (2020). Macroeconomic Effects of Uncertainty: A Big Data Analysis for India. RBI Working Paper 4.

Richardson A. van Florenstein Mulder T. and Vehbi T. (2019). Nowcasting New Zealand GDP using machine learning algorithms. IFC Bulletin 50.

Sawaengsuksant P (2019). Standardised approach in developing economic indicators using internet searching applications. IFC Bulletin 50.

Siksamat S. (2021). Collecting data: new information sources. IFC Bulletin 54.

Tissot B. (2017). Big data and central banking", IFC Bulletin 44.

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

Wanitthanankun J. & Dummee J. (2017). Micro Data Usage Enhancement in Bank of Thailand. MyStats, Department of Statistic Malaysia.

Washimi K. (2021). Venture Capital and Startup Innovation-- Big Data Analysis of Patent Data. BOJ reports & research papers, Bank of Japan.

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

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; operationalize 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 & 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 & 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 JGB Market (currently available only in Japanese)	Collected from financial institutions located in Japan	Analyze the structure of the Japanese repo market	R
	Release of "Statistics on Securities Financing Transactions in Japan"	Collected from financial institutions located in Japan	Widely share with statistics users 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 the productivity growth, and other	R
	Analysis of business and consumer sentiments	Economy Watchers Survey	Analyse business and consumer sentiments using comments from respondents of the survey by text analysis	R, Python

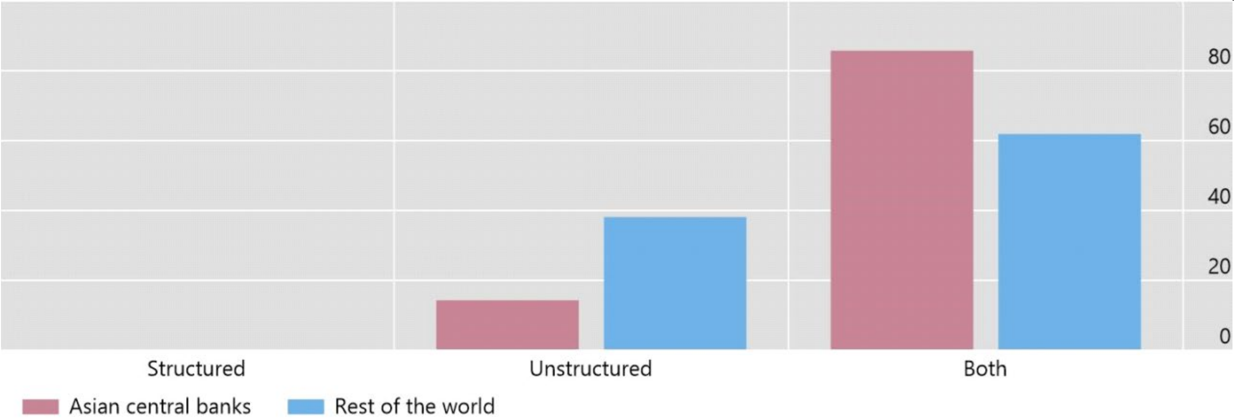
Bank of Thailand	Leading indicators for export	Thai Customs Department	Develop leading indicators	Python
	Manufacture sector structure	Manufacturing firm census from NSO	Understand the structure of manufacturing sectors	Stata
	High-rise Residential property occupancy rate	Electricity bills, Provincial Electricity Authority	Monitor real demand for high-rise property	RStudio
	Use internet search technology	Google Trend/Correlate	Develop indicators to help monitor economic condition	Google Trend/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 custom and Bank of Thailand	Develop indicator for monitor Thai export value	Python
	Stylized facts on invoicing currency and natural hedge of Thai exporter	Thai custom and Bank of Thailand	Explore invoice structure and natural hedge of Thai exporter	RStudio
	Self-employ labor income	Labor Force Survey from NSO	Determine self-employ 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 trades 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 relationship between entities and provide search capability.	Python, Django, ElasticSearch3, 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).

Central bank definitions of big data and main sources

As a percentage of respondents

Figure 1



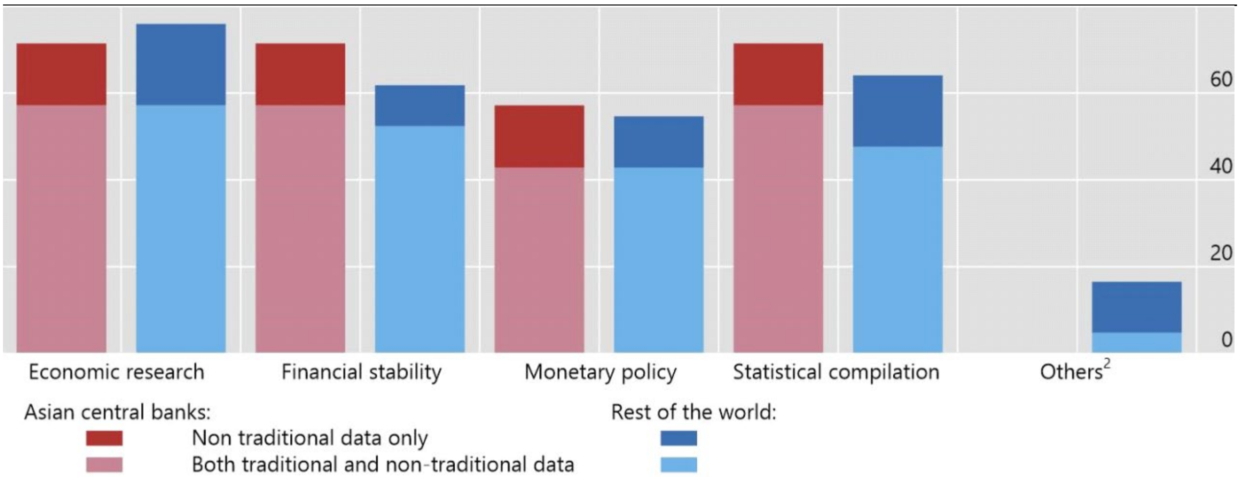
The sample includes 7 Asian central banks and 43 non-Asian central banks. Respondents could select multiple options. Non-traditional data include “unstructured data sets that require new tools to clean and prepare”, “data sets with a large number of observations in the time series”, “data sets that have 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.

Purposes for which central banks use big data[1]

In per cent

Figure 2



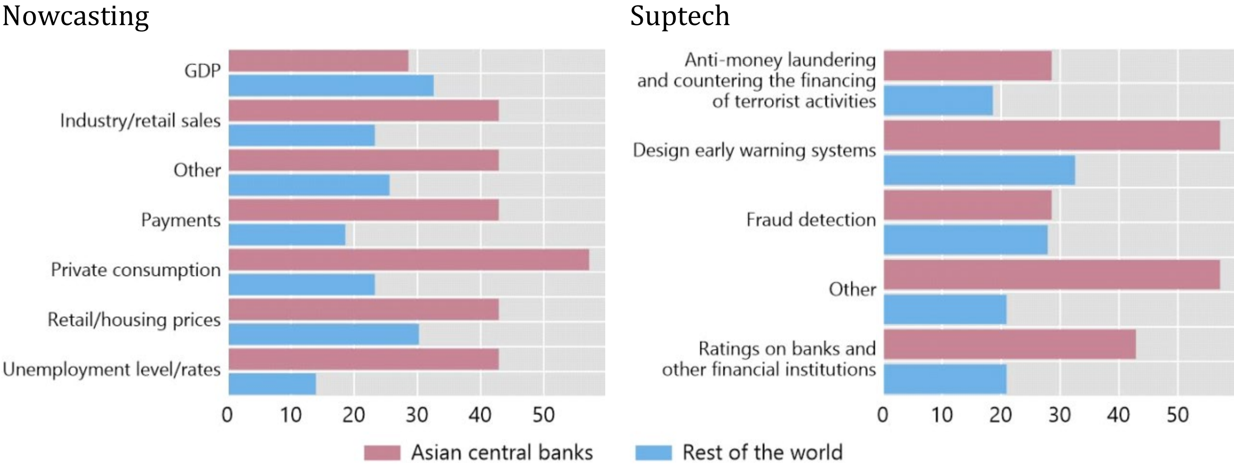
[1] Respondents could select multiple options. See footnote to Figure 1 for details on how institutions define big data. The sample includes 7 Asian central banks and 42 non-Asian central banks. [2] Includes “monitoring crypto assets”, “cyber security” and “network analysis”.

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

For what specific purposes does your institution use big data?

In per cent of respondents

Figure 3



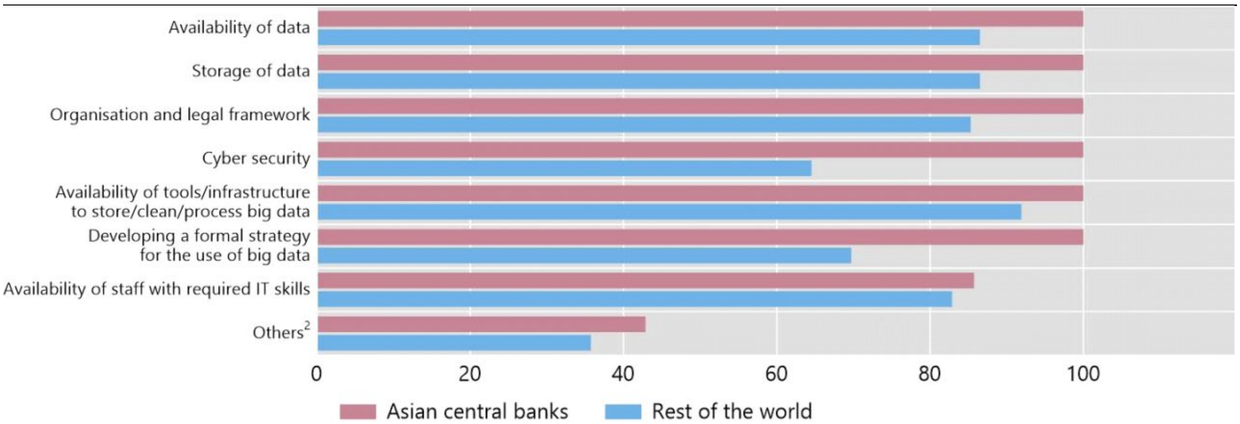
The sample includes 7 Asian central banks and 42 non-Asian central banks.

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

What is the focus of the discussions on big data within your institution?[1]

In per cent

Figure 4



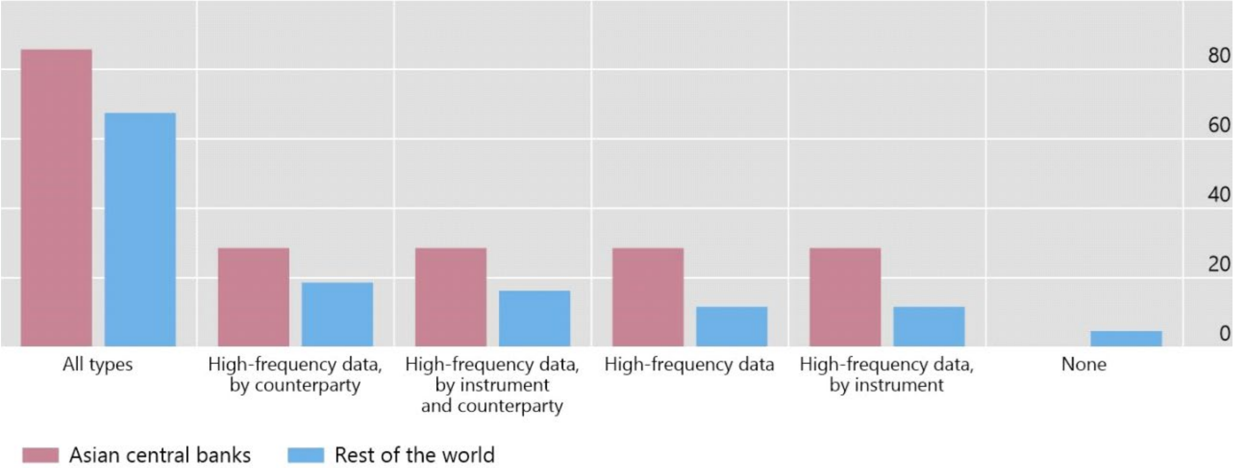
[1] Respondents could select multiple options. The sample includes 7 Asian central banks and 43 non-Asian central banks. [2] Includes "data quality and reliability", "data interpretation" and "data governance".

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

Which types of payments data are useful for your institution?

In per cent

Figure 5



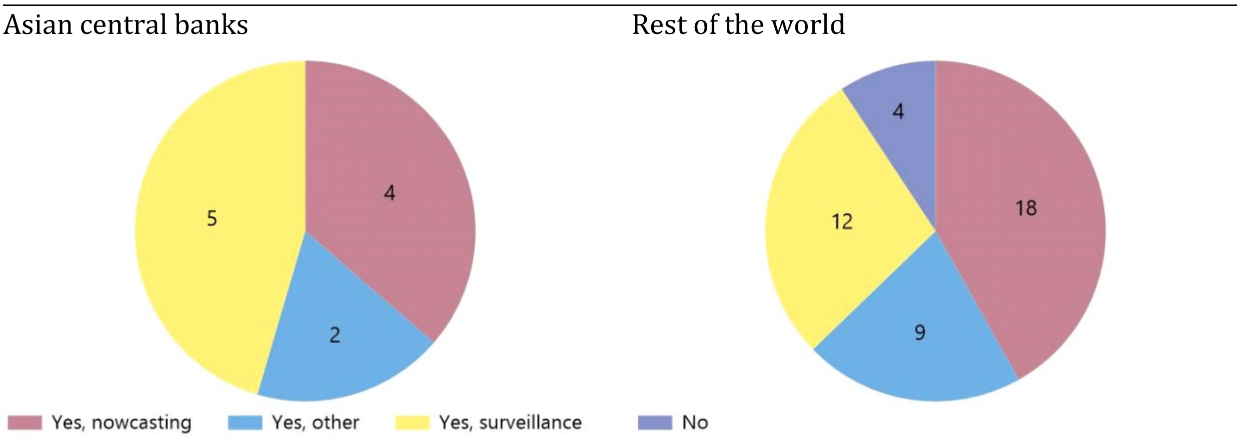
The sample includes 7 Asian central banks and 43 non-Asian central banks.

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

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

Number

Figure 6



The sample includes 7 Asian central banks and 43 non-Asian central banks.

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

1 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 answered the survey. More information on the survey is contained in IFC (2021a) and Doerr et al. (2021).

2 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 2019a).

3 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 perceive as negative. This will bias any analysis based on them.

4 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.

5 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.

6 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.

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