How do central banks use big data and machine learning?

By Sebastian Doerr, Leonardo Gambacorta and Jose Maria Serena*
Bank for International Settlements

Big data and machine learning tools are making rapid inroads in the central bank toolkit. This policy brief examines how central banks define and use big data, leveraging on a survey conducted in 2020 among the members of the Irving Fischer Committee. A majority of central banks use structured and unstructured big data to support their economic analyses and policy decisions, including in the areas of economic research, financial stability and monetary policy. Major challenges include setting up the necessary IT infrastructure and training existing or hiring new staff to work on big data related issues. Further issues include data quality and legal aspects around privacy and confidentiality. Cooperation among public authorities could relax some of these constraints.


The "age of big data" is upending models of data collection and has furthered central banks' interest in the analysis of non-traditional data. In a recent Irving Fisher Committee (IFC) survey, more than 80% of central banks report that they use big data, up from 30% five years ago (IFC (2021)). The high interest is also reflected in the rising number of speeches by central bankers that mention big data or related terms and do so in an increasingly positive light (Figure 1).

Combined with novel techniques such as machine learning or natural language processing, the use of big data in central banks offers new opportunities to understand the economy and financial system better and to inform policy. Yet, the advent of big data also poses new challenges. These range from operational concerns about building the relevant infrastructure and hiring new staff to obstacles of data quality and considerations about data privacy and confidentiality.

In this policy brief, we examine how central banks define and use big data, as well as the opportunities they see and challenges they face, summarising the main insights of recent research (Doerr et al (2021)). In doing so, we leverage on a survey conducted in 2020 among the institutional members of the Irving Fischer Committee of the Bank for International Settlements.

Figure 1: Central banks' interest in big data is mounting
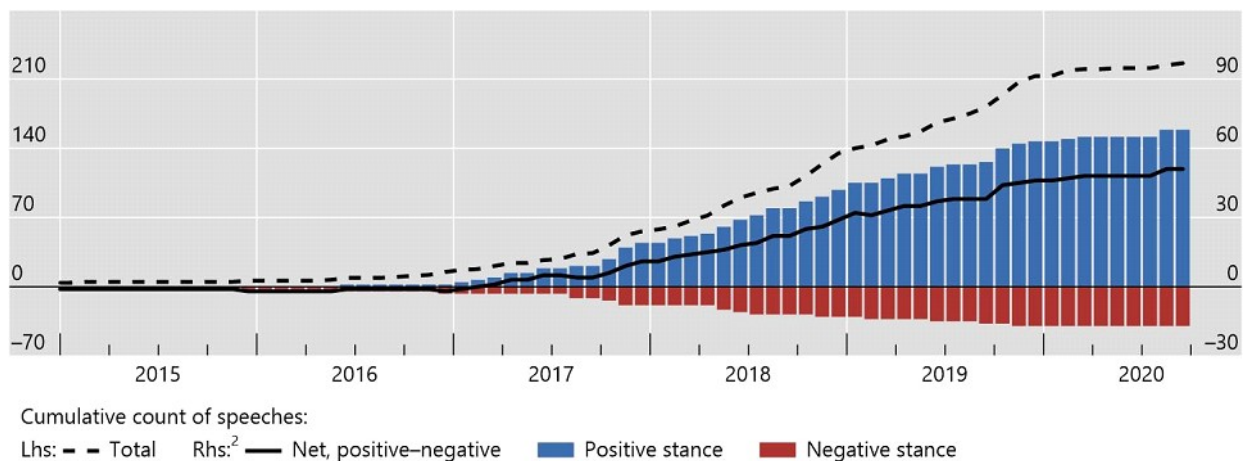
Number of speeches mentioning keywords "artificial intelligence", "big data" or "machine learning". Source: Doerr et al (2021).
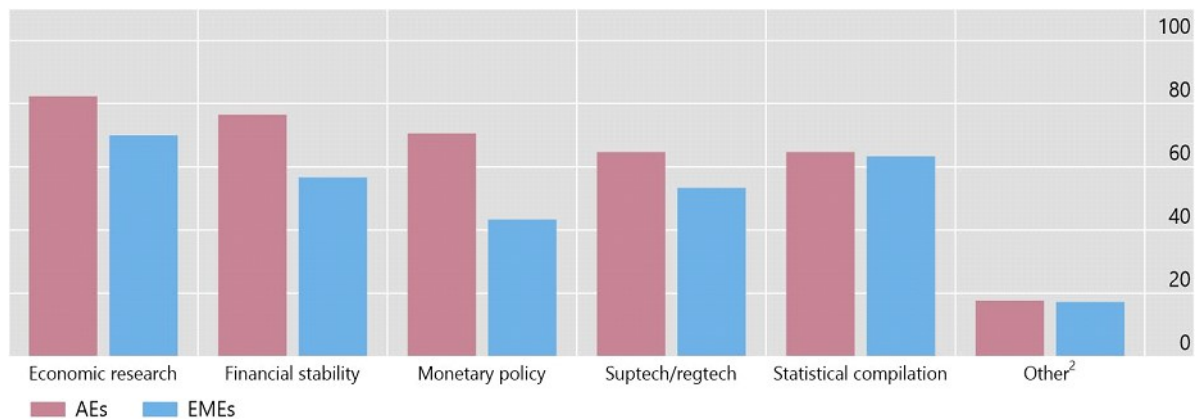
Use cases

Central banks define big data in a comprehensive way. Over 90% of respondents report that big data includes unstructured datasets that require new techniques to analyse. In addition, 80% of respondents report that datasets with a large number of observations in the time series or cross-section constitute big data. Around 75% of respondent refer to big data as datasets that have not been part of their traditional data pool. Finally, around two-thirds of respondents consider large structured databases as big data.

The variety in definitions is also reflected in the multitude of sources that are currently used for data analysis in central banks. These range from structured administrative data sets such as credit registries to unstructured non-traditional data obtained from eg newspapers, mobile phones, web-scraped property prices, or social media posts.

Big data and machine learning tools are making rapid inroads in the central bank toolkit: more than 80% of central banks routinely use big data and related techniques to support their emerging business needs. Central banks use big data in a variety of areas, including economic research, financial stability and monetary policy. Central banks of advanced economies use big data in these areas relatively more than emerging economies (Figure 2). Big data are also used for supervision and regulation (suptech and regtech applications) and statistical compilation, largely to a similar extent in advanced and emerging market economies.

Figure 2: For what purposes do central banks use big data?

Share of respondents that selected each respective answer to the question "For what general purposes does your institution use big data?". Source: Doerr et al (2021).

Four specific use cases are particularly widespread. First, a large and increasing number of central banks support their economic analyses with nowcasting models using big data. These models produce high-frequency forecasts that can be updated in almost real-time. Nowcasting models often rely on a variety of data sources, ranging from structured information to datasets retrieved from non-traditional sources.

Second, central banks use big data techniques to measure other aspects of economic activity. For example, they use natural language processing to produce economic or policy uncertainty indices from textual data.

Third, central banks are exploiting financial big data to support their financial stability analysis (Cœuré 2017). These data, which encompass the large proprietary and structured datasets compiled within central banks, include trade repositories with derivative transactions, credit registries with information on loans, and individual payment transactions. Trade repositories have helped, for instance, to identify networks of derivative exposures.

Finally, around 60% of central banks report the use of big data for suptech and regtech applications. Most applications are yet exploratory, but could become more widespread (IFC 2021). Specifically, natural language processing is used to augment traditional credit scoring models, drawing on information from news media or financial statements. Big data algorithms are incipiently used for anti-fraud detection, identifying suspicious payment transactions, and fraudulent loan clauses. Central banks also support the use of regtech applications by financial institutions to meet compliance requirements.

Challenges

The use of big data in central banks provides not only opportunities, but also new challenges Figure 3 illustrates the main topics of discussion within central banks. Setting up the necessary IT infrastructure constitutes a key

challenge, as it entails high fixed costs. Providing adequate computing power and finding the right software tools also prove challenging. Further, there are difficulties in training existing or hiring new staff to work on big data related issues, a problem compounded by an acute shortage of data scientists, who are also much sought-after in the private sector (Cœuré (2020)).

Figure 3: What is the focus of the discussions on big data within institutions?
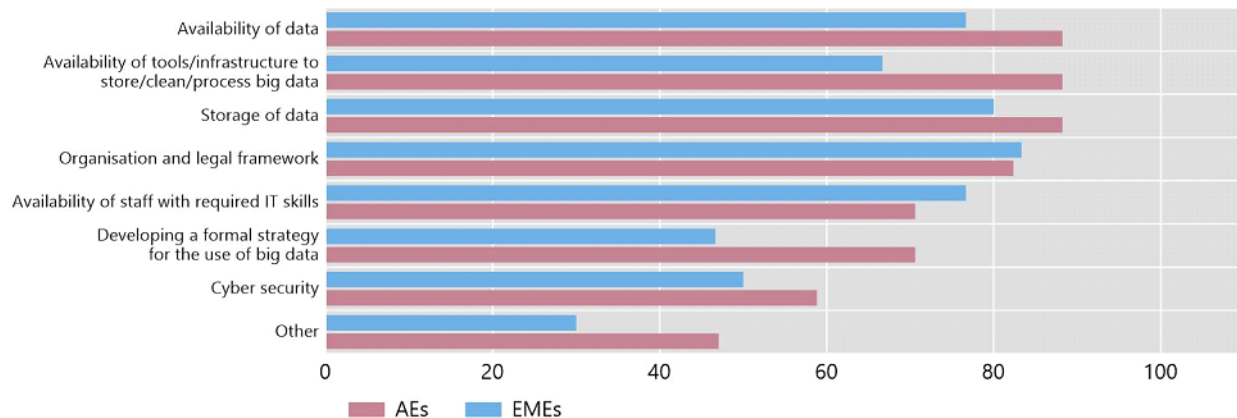
Share of respondents that selected each respective answer to the question "What is the focus of the discussions on big data within your institution?". Source: Doerr et al (2021).

The legal underpinnings of the use of private and confidential data present another challenge, as the world of data is transforming rapidly. Traditionally, most data were collected and hosted within public institutions and hence readily available for analysis. Over the last years, data creation has migrated to the private sector. Companies now own vast troves of confidential and granular data that resides outside the direct reach of central banks or regulators. Considerations about ethics and privacy come with the use of potentially sensitive information that rests in the hands of private companies. For example, are central banks allowed to use data on users' search history or web-scraped social media postings in their analyses?

Fundamentally, these considerations reflect that citizens value their privacy and might be unwilling to share their data if given a choice. Privacy concerns vary across and within jurisdictions: for example, women and older cohorts are less willing to share their financial data; and so are individuals in richer countries (Chen et al (2021)). To varying degrees, central banks and other policy makers need to convince the public that data will not be used for unjustified or unauthorized infringement on individuals' right to privacy.

Central banks are also concerned about the quality of big data: non-traditional data is often a by-product of economic or social activity and not curated for the purpose of further analysis; financial big data could suffer from reporting problems. Data cleaning (eg in the case of media, social media or financial data), sampling and representativeness (eg in the case of web searches or employment websites), and matching new data to existing sources are major challenges to the usefulness of big data.

Policy considerations

Cooperation among public authorities could relax the constraints on collecting, storing and analysing big data. Smaller jurisdictions can particularly benefit from cooperation, as they do not profit from economies of scale when investing in hardware and software. Instead, they could share in the setup costs or "rent" the necessary storage capacity and computing power, as well as staff resources, from larger jurisdictions.

Examples of successful cooperation include the ECB's AnaCredit database, which collects harmonised data from euro area member states in a single database to support decision-making in monetary policy and macroprudential supervision. Likewise, the BIS collects and processes confidential banking data in cooperation with central banks and other national authorities, including the International Banking Statistics, and data collected in the International Data Hub.

That said legal obligations to store the raw data within national boundaries could restrict the sharing of confidential data. There would be a need to agree on, say, cloud computing contracts, as well as on rules for data use and the protection of confidentiality. Protocols to ensure algorithmic fairness would have to be developed and ratified. ◼

References

Chen, Sharon, Sebastian Doerr, Jon Frost, Leonardo Gambacorta and Hyun Song Shin (2021): "The fintech gender gap", BIS Working Paper No. 931.

Cœuré, Benoit (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, 24 November.

Cœuré, Benoit (2020): "Leveraging technology to support supervision: challenges and collaborative solutions", speech at the Financial Statement event series, Peterson Institute, 19 August.

Doerr, Sebastian, Leonardo Gambacorta and Jose Maria Serena (2021): "Big data and machine learning in central banking", BIS Working Paper No. 930.

Irving Fisher Committee (2021): "Use of big data sources and applications at central banks", IFC Report no 13.

About the authors

Sebastian Doerr *is an economist in the Monetary and Economics Department at the Bank for International Settlements. He holds a PhD in economics from the University of Zurich. His research focuses on the consequences of financial innovation for financial stability and consumer welfare, as well as on the impact of macroeconomic trends on the financial sector.*

Leonardo Gambacorta *is the Head of the Innovation and the Digital Economy unit at the Bank for International Settlements. His main interests include the monetary transmission mechanisms, the effectiveness of macroprudential policies on systemic risk, and the effects of technological innovation on financial intermediation.*

Jose Maria Serena *is an economist in the Monetary and Economics Department at the Bank for International Settlements, and previously worked at the Bank of Spain. He develops research and analysis on financial intermediation, with a particular focus on non-bank financial institutions, and various corporate finance issues.*