Irving Fisher Committee on Central Bank Statistics

BIS

Uses of AI in statistics: the central banks' experience

Bruno TISSOT

Head of Statistics and Research Support, BIS

Head of the Secretariat of the Irving Fisher Committee on Central Bank Statistics (IFC)

11th meeting of the Arab Statistics Initiative, 13 November 2024

The views expressed are those of the author and do not necessarily reflect those of the BIS or the IFC.

1. Data issues and the BIS IFC Committee of central banks
2. The information revolution: new data and tools
3. Statistical production
4. Forecasting exercises
5. Challenges with big data and innovative tools

1. The IFC: BIS Statistics Committee of central banks

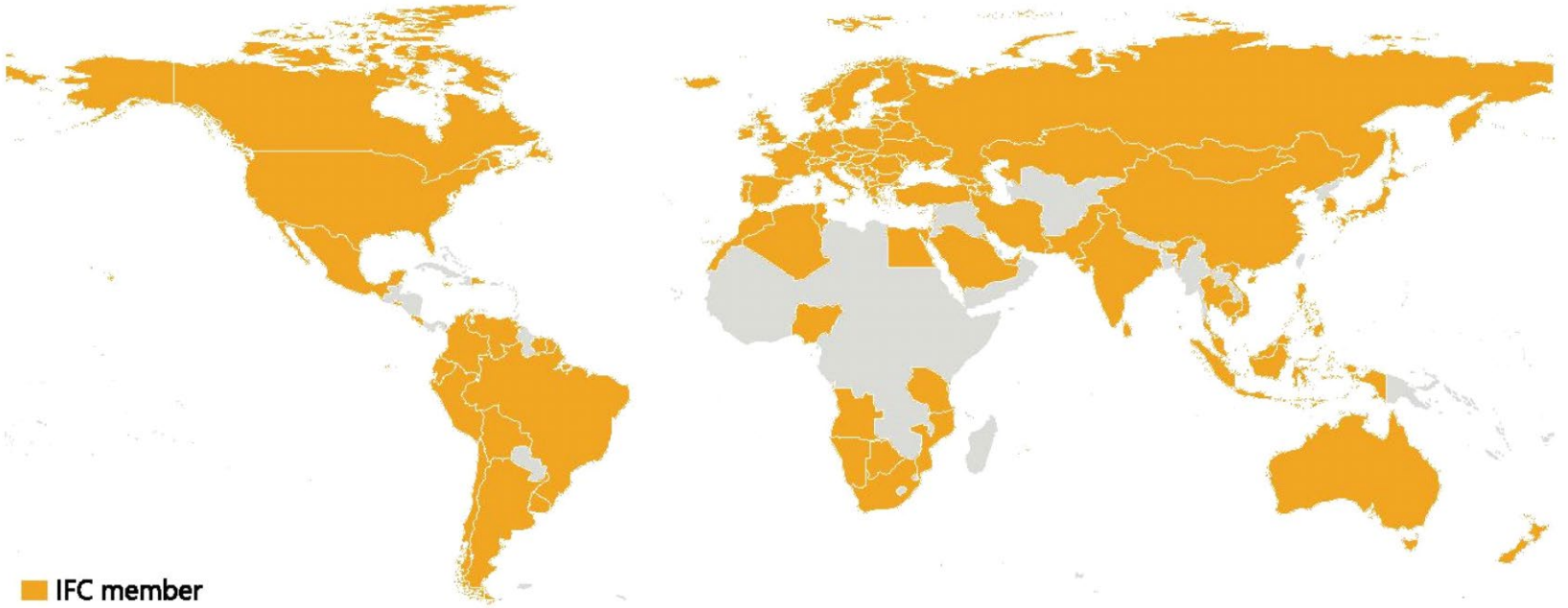
- BIS Irving Fisher Committee on Central Bank Statistics (IFC)
  - ✓ **Forum** of central bank economists and statisticians
  - ✓ Focus on **statistical issues** of interest to central banks
  - ✓ Governed by the international central banking community

  →*reporting to the **BIS All Governors'** Meeting*
  - ✓ **Chair**: Alberto Naudon, Board Member of the Central Bank of Chile
  - ✓ **Sharing of experience and knowledge centre**
- Importance of **collaboration and partnership** with other organisations, including central banks but also international organisations, NSOs, academia and the private sector … especially in the area of big data and AI

1. IFC (cont'd): Membership

- **107** IFC members

Members of the Irving Fisher Committee (IFC) on Central Bank Statistics



The use of this map does not constitute, and should not be construed as constituting, an expression of a position by the BIS regarding the legal status of, or sovereignty of any territory or its authorities, to the delimitation of international frontiers and boundaries and/or to the name and designation of any territory, city or area.

Source: BIS.

1. IFC (cont'd): Established work on data innovation topics (last decade)

- *Artificial intelligence in central banking, BIS Bulletin, # 84*
- *Data science in central banking: applications and tools , IFC Bulletin #59*
- *Machine learning in central banking, IFC Bulletin, #57*
- *Issues in data governance, IFC Bulletin, #54*
- *IFC Report #15 How central banks communicate on official statistics*
- *IFC Report #13 Use of big data sources and applications at central banks*
- *IFC Report #11 Computing platforms for big data analytics and artificial intelligence*
- *IFC Report #9 Business intelligence systems and central bank statistics*
- *The use of big data analytics and AI in central banking, IFC Bulletin #50*
- *Big Data, IFC Bulletin, #44*

1. IFC (cont'd): … based on experience reported by central banks

Selected list of central bank use cases of machine learning — Table A1

Main method	Application type			
	Information collection	Macro/financial analysis for monetary policy	Payments oversight	Supervision
Tree-based methods	Banco de Portugal, Bank of Israel, Deutsche Bundesbank, ECB, Magyar Nemzeti Bank	Bank Indonesia, Bank of France, Reserve Bank of Australia, Reserve Bank of New Zealand	Central Bank of Iceland	Bank of France, Bank of Italy, Bank of Japan, Banco de Portugal, Bank of Spain, Central Bank of Brazil
Neural networks	ECB	Bangko Sentral ng Pilipinas, Bank Indonesia, Bank of Canada, Bank of Korea, Central Bank of Chile, Central Bank of Malaysia, Bank of Canada, Bank of England, Deutsche Bundesbank, ECB	Bank of Canada, Bank of Italy, Bank of Thailand, Central Bank of Ecuador, De Nederlandsche Bank	Bank of France, Bank of Greece, Central Bank of Brazil, Deutsche Bundesbank, Hong Kong Monetary Authority
Large language models	Deutsche Bundesbank	Bangko Sentral ng Pilipinas, Bank Indonesia, Bank of Korea, Deutsche Bundesbank, Federal Reserve	Bank of Korea	Central Bank of Malaysia, ECB, Federal Reserve
Other techniques	De Nederlansche Bank, Deutsche Bundesbank	Bank of Italy, Czech National Bank, South African Reserve Bank		Bank of Canada, Bank of Slovenia, Bank of Spain, Bank of Thailand, Central Bank of the Republic of Austria[1], ECB[1], Federal Reserve, Monetary Authority of Singapore[1]

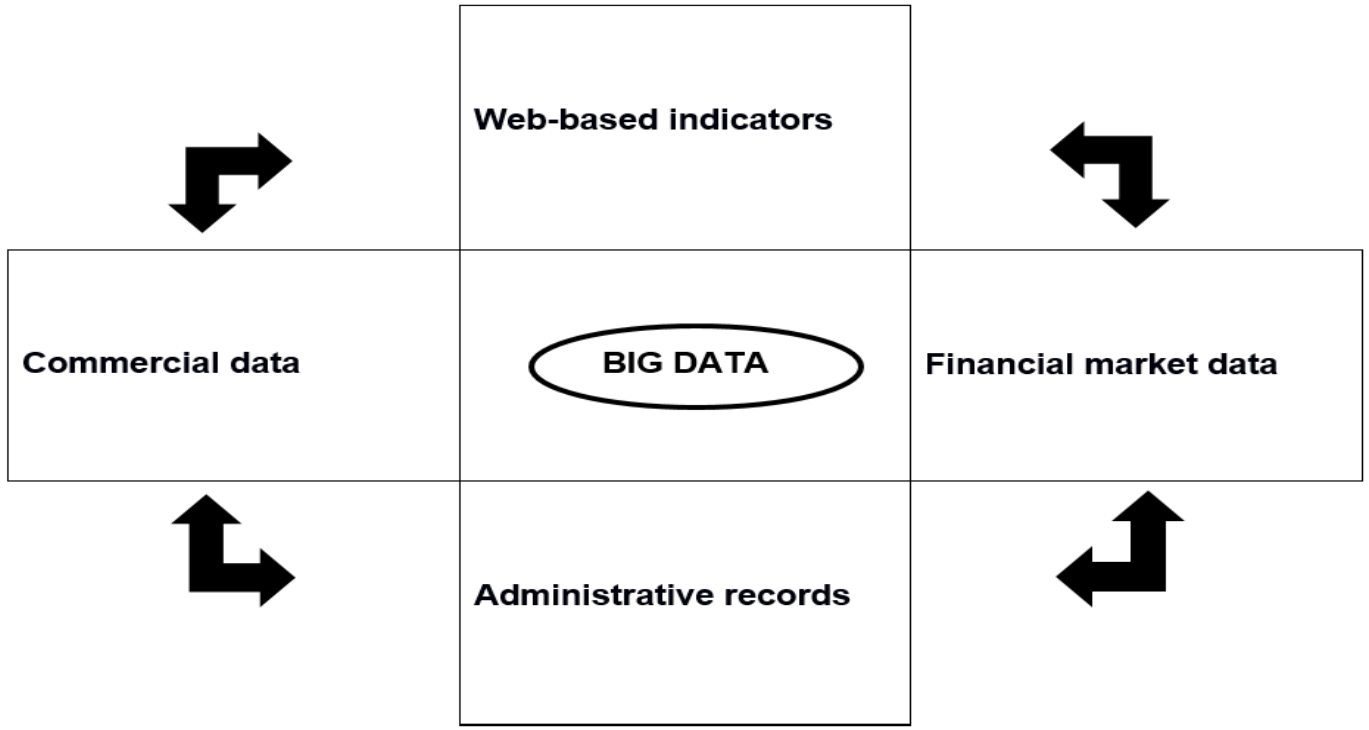
Sources: Araujo et al (2022, 2023); Beerman et al (2021); national central banks; Irving Fisher Committee.

[1] Specific technique not disclosed publicly.

2. The information revolution: new (or not so new?) data and tools

- Big Data
- Big Data Analytics, Artificial Intelligence and Machine Learning
- Generative AI
- Data Science

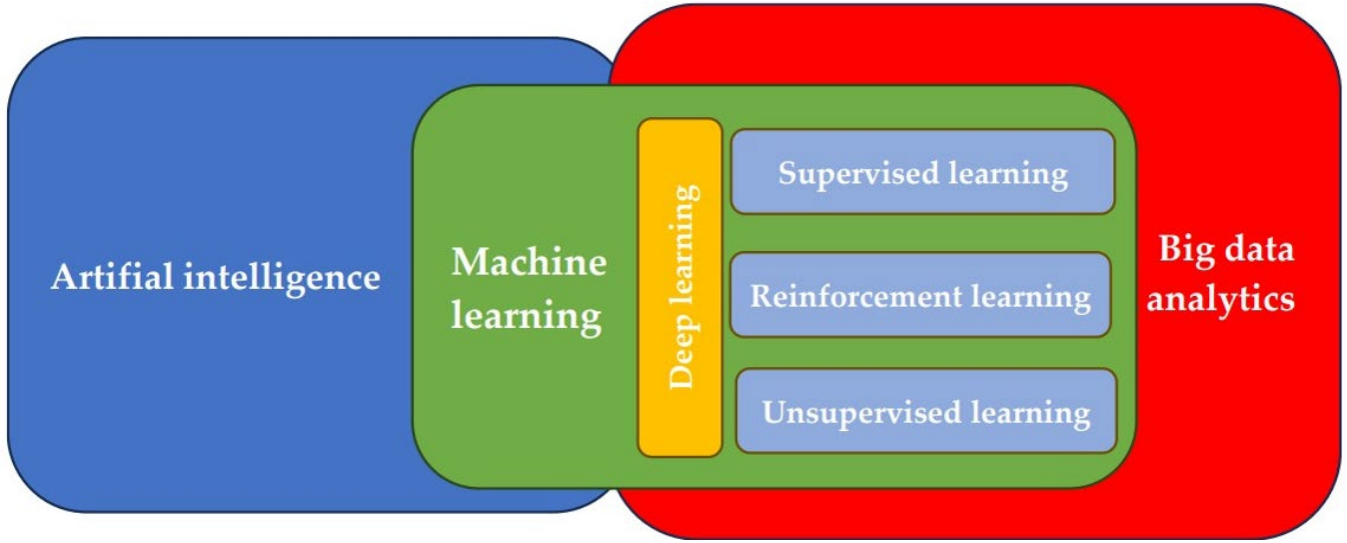
2. Information revolution (cont'd): 4 types of "Financial Big Data"…



Source: IFC.

- New data reflecting the **advance of technology** eg impact of digitalisation, surge in geospatial data, mobile phones' mobility trends reports…
- … and not so new data reflecting innovative ways to **make use of existing structured or unstructured information** (eg text, organic administrative datasets)

2. The information revolution (cont'd): … new techniques…



Source: FSB.

- **Big Data Analytics** – general analysis of big data sets (large and/or complex)
- **Artificial Intelligence (AI)** – theory and development of computer systems to perform tasks that traditionally required human intelligence
- **Machine Learning (ML**; subset of AI techniques) – method of designing a sequence of actions to solve a problem that optimise automatically through experience and with limited or no human intervention

2. The information revolution (cont'd): … including Generative AI

- Models learning the patterns and structure of input information ("training data") and **generating new but similar data** without much user expertise

- **Increased interest for textual information**

- **Natural Language Processing** (NLP): turning *unstructured* text information into *structured* data

- **Large Language Models** (LLMs) eg ChatGPT: neural networks that are trained to predict the next word in a given sequence of text

*→Key role of **transformers** in capturing the relationship between different components of a text – even if they are far apart*

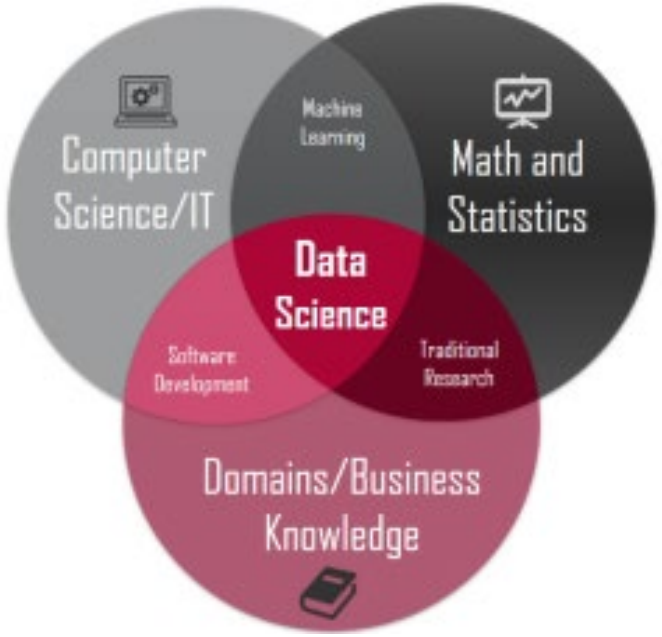
- **Multiple use cases**

→Data dissemination, communication and information search

→Production of analyses (statistical releases)

→Quality control, detection of patterns in the data…

2. Information revolution: need for a broad data science approach…



Source: N Cheng, "Data analyst primer: the essential guide", Medium, 27 Aug 2020, after D Conway, "The data science Venn diagram" September 2010.

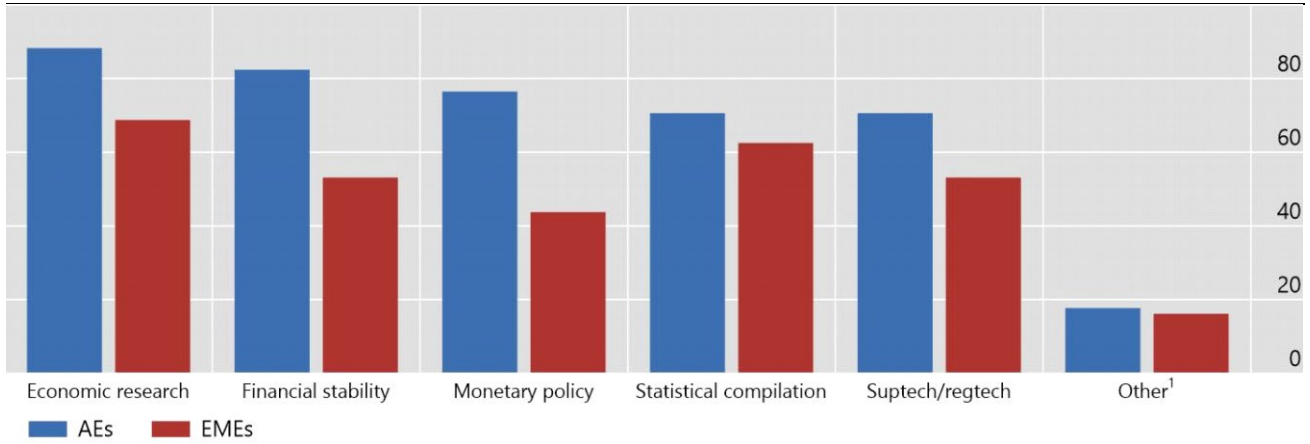
- **Data Science**: study of data (including techniques for extracting insights)
- **Interdisciplinary field** that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms
- **Intersection** of (1) information technology; (2) mathematical and statistical methods; and (3) business, or "subject-matter" expertise

2. … already supporting ongoing projects in central banks…

- **Big data projects** already support most of central banks' main functions…

For what general purposes does your institution use big data?

In per cent of respondents



[1] Includes monitoring crypto-assets, cyber security, impact and network analysis.

Source: IFC Report no 13 (2021).

- **… in a constantly evolving & uncertain landscape for data and tools** (eg quantum / edge computing, explainable AI, Blockchain / DLT, synthetic data etc.

3. Statistical production

- Support of **current official statistical production of economic indicators**
  - Typically a task for NSOs… but not only…
  - … esp. for central banks willing "complementary" information

- Big data / new tools can **help statistical compilers** by
  - Facilitating **data collection** by using readily available information (web scraping) and improving statistical agility to address unexpected data needs
  - Reducing **reporting burden**
  - Alleviating the **cost** of conducting traditional surveys (eg sending of dedicated staff versus capturing data on the web)
  - Improving **timeliness** (with new data sources available more rapidly)
  - Computing **advanced estimates** of "hard" but lagging indicators
  - And possibly being more **accurate (?)** with "organic", observed data (vs reported data)

3. Statistical production

- Support the computation of **additional indicators**
  - **Higher frequency** estimates: daily CPI, weekly GDP
  - More specific and **granular information**
    - Scraping of a wide range of **detailed** data
    - Measurement at **different points of time**
    - Various **sub-items** of interest eg "core" CPI, property prices, unemployment
  - **Substitute ("buffer") information** when official data are missing
    - When the official statistical production chain is **disrupted** (eg Covid-19 lockdowns)
    - Comprehensive statistics may not be fully available in **developing statistical systems**

3. Statistical production

- Provision of **new insights on economic phenomena**
  - **Distribution** information
    - Generally missing in the SNA framework
    - Situation can vary across populations / firms
  - **Enhanced micro-macro** integration
    - Aggregates compiled from granular "Lego bricks" that can be informative
    - Easier production of sub-items depending on circumstances (eg sectoral shocks)
  - **New types of indicators eg**
    - Asset prices (eg properties) for instance by using web scraping
    - Soft indicators (eg expectations, uncertainty) eg using NLP tools

4. Forecasting exercises

- **AI support for various types of forecasting exercises**
  - Use of **real-time** indicators for forecasting the present / the past
    - → Nowcasting, advanced estimates
  - Enhanced **short-term projections**
    - → Better lead time
    - → Updating possibilities as incoming data (constantly) arrive
  - **Complement "traditional" forecasting tools** by detecting
    - → Non-linearities
    - → Abnormal patterns
    - → Turning points

4. Forecasting exercises

- **Making use of more information for forecasting**
  - Large **variety of big data sources**

    →Web-based eg Google Trends, internet sales

    →Financial market indicators

    →Qualitative (textual) information eg in social medias

  - **Selection** of explanatory variables

→**Greater choice**, more potential data can be considered ex ante

→ Possibility to "**zoom in**": focus on more granular items

→Flexibility for **adjusting forecasts** depending on

- the time horizon eg short/long-term
- the actual date of the exercise

  - **Use of ML algorithms** to both deal with & select incoming data

5. Many challenges posed by big data and innovative tools…

… hampering a widespread use of big data – resources, IT, staff, ethical & legal aspects, security & cyber risk, privacy & confidentiality, external dependencies, data sharing/access, etc. etc…

Big data challenges

Word count on challenges

[1] Open answers are transformed by removing special characters, white spaces or stop words (such as "the", "a" or "we"). A text-mining algorithm then counts the frequency of individual words. Words mentioned more frequently appear larger.

Sources: IFC Report no 13 (2021); authors' calculations.

5. Especially for supporting policies

- **Challenges reflecting**
  - The variety of big data sets and tools
  - Their complexity
  - Time dependency aspects

- **Associated policy issues when using Big Data/AI**
  - Accuracy
  - Reputational risk
  - Altering decision-making

Thank you!!

Questions?

bruno.tissot@bis.org
IFC.secretariat@bis.org

Selected Publications & References

available at https://www.bis.org/ifc/publications.htm?m=3%7C46%7C94