15 December 2023

Guidance Note 4 – Central banks' use of time series products

This guidance note[1] is based on an internal survey of central banks (joined by a few international institutions and statistical agencies) organised through the network of the Irving Fisher Committee on Central Bank Statistics (IFC).[2]

It sheds light on five main aspects related to central banks' use of time series products, namely:

- *whether time series data are treated separately from other data types;*
- *the reported use of time series software;*
- *the relative strengths and shortcomings of alternative approaches;*
- *the potential need for a new software generation for dealing with time series data in the "big data" era; and*
- *some lessons looking forward.*

1. Dealing with time series data

Central banks have various ways in setting up an end-to-end process for their products to deal with time series, from getting data "in the building" to producing analyses and publications.

One particular issue is whether time series data, which can be defined as a sequence of data points indexed over a time interval, should be treated as distinct from other forms of data (especially in today's new era of granular and unstructured data).

Indeed, a main lesson of the IFC survey is that **most respondents treat time series as distinct from other types of data** in at least some key aspects of data management and access (Graph 1).

2. Reported use of time series products in central banks

There is a **variety of commercial and/or proprietary systems** that can be used to support the processes dealing with time series. This is particularly the case for *time series databases*, ie those

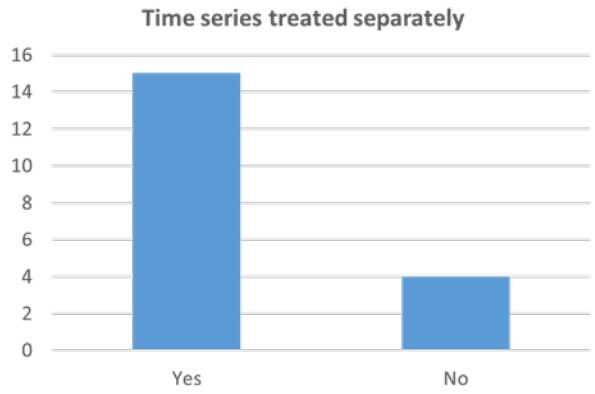
[1] This analysis produced by the staff of the Bank of England (BoE) (ryland.thomas@bankofengland.co.uk) and the IFC Secretariat (bruno.tissot@bis.org; ifc.secretariat@bis.org) does not necessarily reflect the views of the BoE, the BIS or IFC members and the various institutions mentioned in this note. Bilyana Bogdanova, Alessandro Bonara, Robert Kirchner, Edward Lambe, Rafael Schmidt, Olivier Sirello and Xavier Sosnovsky are very much thanked for their comments and suggestions.

[2] Cf Thomas, R (2023): "Future of time series: preliminary results from a BIS-IFC survey of central banks and statistical agencies", presentation at the 3rd IFC workshop on "Data science in central banking", October 2023, Rome; the presentation can be downloaded here.

software systems used for storing and serving time series. In this context, institutions engaged in statistical production have had access to a broad array of proprietary tools.[3]

Time series treatment

Graph 1



Source: IFC internal survey on the use of time series products (2023).

The IFC survey shows that **one of the products most frequently used by the central banking community in recent years has been FAME** (Forecasting Analysis and Modeling Environment).[4] About **three fourths of the survey's respondents have been relying on this product in the past or are still continuing to do so**. Publicly reported examples include the Central Bank of Spain, whose Statistics Department had been using this product across the different stages of its statistical workflow (ie compilation, dissemination and transmission to international organisations).[5] The ECB did follow a similar approach for many years for the production/ dissemination of time series and macroeconomic statistics.[6]

There are also other software available in the market place, such as Matlab, STATA, Eviews, etc. For instance, one product used particularly by national statistical offices (NSOs) in recent decades has been SAS (Statistical Analysis System).[7] Publicly mentioned examples of SAS use cases have

[3] Proprietary products relate here to software owned and sold by a commercial provider and that, as "closed source software", cannot be changed directly by users.

[4] FAME is a time series database released in 1981 and owned by FIS (Fidelity National Information Services, Inc.) Global. FAME, is the data storage/analytical tool provided as part of the Market Data Suite (formerly MarketMap Analytical Platform or MAP) product in the FIS Integrated Statistical Data Production System.

For its use in the context of official statistics, see Kumthekar, R and P Bakos (2019): "Integrated statistical production system with Generic Statistical Business Process Model (GSBPM)", September; and Hinge, D (2021): "Data Analytics Partner: FIS", *Central Banking*, June

[5] Cf Bollo, E (2017): "SDMX system in the Banco de España", presentation at the Eurofiling XBRL Week, Frankfurt, June.

[6] Cf Witt, E and J Blaschke (2019): "ECB data for analysis and decision-making: data governance and technology", *IFC Bulletin*, no 49.

[7] A statistical software suite developed by SAS Institute for data management, advanced analytics etc.

included Statistics Canada,[8] the Irish Central Statistics Office, INSEE in France[9] and Statistics South Africa.[10]

However, a recent but important development in official statistics has been the increased interest for alternative open sources software, such as R or Python. And, indeed, the IFC survey shows that **several central bank respondents are actively considering transitioning away** from their traditional set ups.[11] A similar feature has been observed in the NSO community; for instance the Irish NSO recently decided "*to replace its current extensive libraries of SAS code with R code over a 5-year period and be fully operational in an R environment across its statistical functions by the end of 2027*".[12]

Another key insight of the IFC survey is that a growing part of central banks use bespoke systems. **This suggests that the central banking community has so far identified no single, commonly used market product** that would be the preferred option to handle time series in the new big data information era.

3. Relative strengths and shortcomings of alternative approaches

The choice of a particular product to handle time series typically depends on its relative strengths and shortcomings. These can reflect multiple aspects, such as timeliness, reliability, performance, adaptability and ease of upgrading; or costs such as licencing, development and maintenance.

In general, the **main reasons reported for the continued attractiveness of proprietary systems** relate to the following aspects:

- The ability for **dealing with (well structured) time series data in a way that is consistent with the statistical business process models in place** in the institution.[13]
- The availability of **dedicated, adequate and well-functioning functionalities** to handle traditional financial and macroeconomic time series effectively.
- The **stability and efficiency of the products** supported by commercial providers (but requiring nevertheless proper internal support for implementation and maintenance).
- The **degree of embedding** in the analytical systems developed within user central banks, implying that moving away from existing solutions could be costly.

---

8 Which for instance provides the SAS Assistant to help users with little experience to generate successful tables.

9 For a recent overview of the situation in INSEE, see Lair, J-S (2022): "Placer les données au plus près des métiers".

10 In particular its Business Modernisation programme aims to improve data and information management leveraging a SAS platform; cf Statistics South Africa Annual Report 2022/2023.

11 Similar trends were reported on the occasion of the 5th conference on "Non-traditional Data, Machine Learning, and Natural Language Processing in Macroeconomics" organised in November 2023 by the Federal Reserve Board.

12 United Nations Economic and Social Council (2023): "Journey to using R – experience of the Central Statistics Office of Ireland", *ECE/CES/2023/19*, June.

13 For example, it has been argued that the core FAME analytic database management system can contribute to the modernisation of statistical business processes by dealing with well-structured source files based on the Statistical Data and Metadata eXchange (SDMX) information model that is widely used by international organisations, NSOs, and other data-producing agencies to streamline the transmission of data and strengthen their dissemination; cf FIS MarketMap FAME Analytic Platform (2018): "*Modernised statistical business process models for public institutions*".

- **Security features and the degree of compliance with regulatory standards**. These aspects are crucial when handling sensitive financial and economic information (in particular confidential granular data) and can make proprietary solutions more appealing, as they often come with robust security protocols, regular updates and patching to close known vulnerabilities, and effective regulatory compliance.

There are however a number of important factors driving central banks' attention to explore new products for dealing with time series data:

- **Business needs have been evolving towards end-to-end integrated solutions** encompassing open source analytical tools[14] and specialised Business Intelligence (BI) tools such as Tableau for data visualisation.[15] In this context, one open question is how well proprietary systems in place in central banks can integrate efficiently with these packages. Another is how they are able to address users' needs for new functionalities, for instance enhanced searchability capabilities and metadata management tasks.[16]
- **Micro and big data are also gaining significant momentum in official statistics, requiring scalable solutions**. As pointed out by Witt and Blaschke (2019; cf footnote 6 above), granular data sets are in increasing demand in the central banking community and are also growingly incorporated into policy decision-making.[17] Hence a key consideration is to have adequate and scalable systems to handle unstructured and/or large data sets.[18]
- The evolving data landscape is also requiring **architectural solutions which ensure high flexibility and modularity**. For instance, many central banks have been adapting their current operational processes to make use of new advanced analytical techniques as part of their statistical production pipelines. This adaptation often includes the integration of concepts like Machine Learning Operations (MLOps),[19] which allow statisticians to streamline and enhance the flexible application of machine learning in their workflows.
- **From an operational standpoint, legacy systems can be risky to maintain** as they tend to rely on limited internal knowledge while the new staff joining the institution may come with a different type of IT expertise. This problem can be reinforced by the fact that, reflecting their commercial nature, the public documentation of proprietary systems can be limited. This may not only prevent the sharing of knowledge within and across institutions but also the harvesting

---

14 See Araujo, D, S Nikoloutsos, R Schmidt and O Sirello (2023): "Central banks as users and providers of open-source software", Box 1, overview of the *IFC Bulletin* on "Data science in central banking: applications and tools", no 59, October.

15 See IFC (2019): "Business intelligence systems and central bank statistics", *IFC Report*, no 9, October.

16 It has been argued in particular that the challenges of managing high-dimensional data objects with complex metadata information calls for going beyond using standard tabular-based approaches, as indeed allowed by new languages such as Python; cf Sirello, O (2023): "Metadata handling: how to make the bridge between standards and data science".

17 See Israël, JM and B Tissot (2021): "Incorporating micro data into macro policy decision-making", *IFC Bulletin*, no 56, April.

18 One example relates to the ECB, which has opted to embark on a "transition phase" and launched in 2019 an important project to replace its legacy platform supporting the statistical production of macroeconomic statistics. The Apache Hadoop technological stack (which is based on open-source software tools and is meeting growing interest in the official statistical community when dealing with large amounts of data and computation processes) has been selected for implementing the new platform and adapt the IT architecture to big data technology; see Delic, A and S Pambianco (2021): "SDMX meets Big Data technologies – ECB implementation", September.

19 See Araujo, D, G Bruno, J Marcucci, R Schmidt and B Tissot (2022): "Machine learning applications in central banking: an overview", *IFC Bulletin*, no 57, November.

of the full benefits of new artificial intelligence (AI) techniques eg large language models (LLMs).[20]

- A related factor driving the exploration of the potential of open source products is central banks' **continuous review of IT costs**. Commercial proprietary systems typically require the purchasing of sometimes costly user licenses. However, such expenses should be compared to those for developing and maintaining a complex data architecture on a stand-alone basis.[21]

4. New approaches for dealing with time series data?

Looking ahead, one critical consideration is the extent to which an institution can further enhance its existing system and the potential impact of needed improvements on its current setup. Or, alternatively, what would be the ideal target if this institution would decide to develop its own system all over again.

Answering these questions in depth would necessitate comprehensive market research to evaluate the diverse options available. Yet, a **preliminary message from the survey is that no single solution appears to have emerged to date in the central banking community**. This appears to be the case for the wide range of aspects involved when dealing with time series.

As regards in particular **data storage aspects**, central banks are actively exploring diverse alternatives to address varying business needs. This exploration includes the traditional SQL data warehouses and relational databases, which can be optimal for highly structured data. It also comprises NoSQL databases, which can be more suitable for data that don't fit the conventional relational table format. Additionally, there is a growing interest in open-source, non-relational distributed databases like Apache HBASE, known for their scalability and distributed computing capabilities. Moreover, some central banks are developing bespoke data lakehouses in-house, combining the features of data lakes and warehouses to create a more flexible and efficient storage solution, with the ability to manage a wide range of data types and structures.

But, although these alternative solutions may arguably be better suited for handling larger and irregular data sets, a major reported challenge is to replicate the convenience, speed and efficiency usually provided by the proprietary systems set up for dealing with time series data. Moreover, new, alternative systems may also require important work to integrate with open source analytical tools. Furthermore, there can be large costs and risks involved in developing completely new solutions "starting from scratch".

In view of these challenges, **one middle-way approach followed by a number of central banks has been to separate tasks**, with some of them based on the traditional systems already developed to benefit from their strong functionalities for managing time series, and other tasks relying on alternative tools that can for instance support analytical and visualisation better. From this perspective, the complementing of existing time series databases systems with open source tools has been highlighted as a potentially efficient, low-cost solution.

---

[20] Reflecting the fact that common AI tools are generally pre-trained on public data and may endogenously favour open-source solutions; see United Nations Economic and Social Council (2023): "Report of the seventy-first plenary session of the Conference of European Statisticians", *ECE/CES/2023/18*, June.

[21] See IFC (2020): "Computing platforms for big data analytics and artificial intelligence", *IFC Report*, no 11, Avril.

5. Lessons looking forward

The IFC survey showed that **an important objective for central bank statisticians is to aim for a "single source of truth"**. This means that the data should ideally be collected once, used multiple times and be stored in one system. In particular, the growing reliance on granular information suggests that an efficient setting would be to try to collect micro-level data at a first stage and then to flexibly aggregate them with fit-for-purpose tools addressing different use cases. Yet what is not clear is how best to scale up and integrate new elements into a comprehensive system for managing times series, especially when starting from a complex IT platform built over many years.

One potential solution from this perspective is to **make a greater use of cloud services**, an option that is increasingly under consideration in the central banking community. It can provide important benefits but also raise significant cost implications and other strategic considerations.[22]

Another key consideration is to **ensure that any new set up is suited for the increasingly broader spectrum of the end users of data series in central banks**. This calls for addressing a wide range of requirements, covering the different data needs of, say, conjunctural analysts, forecasters, researchers or data scientists. The implication is that the respective "use cases" need to be clearly identified and mapped out.

In this context, **identified best practices** include (i) analysing the functionalities that current and potential users value the most, (ii) evaluating whether the product can facilitate demands to cope with increasingly different and complex business case requirements, and finally (iii) planning how to manage the transition for existing users to a new set up.

[22] See IFC (2023): "Data science in central banking: applications and tools", *IFC Bulletin*, no 59, October.