Irving Fisher Committee on
Central Bank Statistics

BIS

The post-pandemic new normal for central bank statistics

Saira Jahangir-Abdoelrahman and Bruno Tissot[1]

Executive summary

Statisticians, including those in central banks, have quickly adapted to the consequences of the Covid-19 pandemic. But, three years after the start of the global crisis, can some normalisation be expected in the official statistical landscape, returning us to the situation prevailing before the crisis? Or, alternatively, should central banks be fundamentally rethinking the way they produce and consume data in a "new normal" state of the world?

The pandemic underlined that data producers have to provide more and more varied types of information to their users, not least by leveraging the wealth of sources available. This calls for the production of more high-frequency and timely indicators as well as for providing new types of indicator (eg sentiment indicators, policy credibility indices) on topics that are still insufficiently covered by existing official statistical frameworks. As key elements of national statistical systems (NSS), central banks have been at the forefront of global initiatives taken to address these issues.

The recent crisis was also a reminder that the statistical landscape has to permanently evolve, for instance to make the most of the opportunities provided by new big data analytics and to reach out to user groups more effectively. Central banks have been well positioned from this perspective, drawing on their unique role as producers and consumers of data to support their policies.

Moreover, recent developments have highlighted the urgency of dealing with a number of long-standing challenges. This calls in particular for the adaptation of statistical frameworks to better reflect structural developments in the global economy, especially as regards globalisation, digitalisation in finance and sustainability aspects such as climate change. Therefore, "central bank statistics have to change as well in order to support and reflect structural changes", as argued by Claudia Buch, Vice-President of the Deutsche Bundesbank.

[1] Respectively, Director Statistics Department, Central Bank of Suriname (sjahangir@cbvs.sr); and Head of Statistics and Research Support, BIS & Head of the Secretariat of the Irving Fischer Committee on Central Bank Statistics (IFC) (Bruno.Tissot@bis.org).

The views expressed here are those of the authors and do not necessarily reflect those of the Central Bank of Suriname, the BIS, the IFC or any of those institutions represented at the conference.

We thank Robert Kirchner, Ko Nakayama, Gloria Peña, Silke Stapel-Weber, Luís Teles Dias and Ulf von Kalckreuth for helpful comments and suggestions.

This suggests that central bank statisticians may be facing a quite heavy agenda to address both the new lessons of the pandemic and the longer-term challenges; in other words, they have to continue "*measuring the past to better understand the present and chart the future*", as emphasised by Claudio Borio, Head of BIS Monetary and Economic Department, in his opening remarks. A number of supporting factors could be helpful in this direction.

A first one is innovation, as the new normal for central banks statisticians is likely to rely heavily on the use of data science to perform their traditional tasks. There is no such thing as a rigid, unchanging statistical framework, and statistical sources and tools have to be continuously refined to fit with evolving challenges – for instance to capture the long-term consequences of the pandemic, assess the sustainability of economic development, and track the global footprint of international firms.

Second, a clear message from the central banking community is that many of the challenges faced during recent crises could be addressed by making a better use of the large amount of micro-level information available in today's modern societies. Here again, technological innovation can open up new perspectives, especially to facilitate the use of data that are available "organically" (ie independently of "designed" traditional statistical compilation exercises such as surveys and censuses) and to spur research using very granular but confidential data sets.

Third, the demand for timely, high-quality and varied statistical data is likely to remain strong, calling for statistical frameworks that can be adapted to meet evolving policy objectives and user needs. Several initiatives have been launched to make the global statistical infrastructure more flexible and efficient, for instance to develop global registers and identifiers and promote data-sharing and the access to new sources of information. Achieving rapid progress on these various fronts is a key priority.

Lastly, while central banks' statisticians have proved responsive and innovative in the face of recent crises, further consolidating their contribution to NSS, their continued and close cooperation with other relevant stakeholders will remain essential to address the current complex challenges and further strengthen the role of official statistics in modern societies.

1. Introduction: a "new normal" for central bank statistics?

During the Covid-19 pandemic, official statistics had to quickly adapt to track new developments (Statistical Journal of the IAOS (2022)). As regards central banks, this global shock had two important immediate consequences (Tissot and De Beer (2021)). First, and reflecting their role as producers of official statistics, they had to proactively respond to the various disruptions observed in the provision of economic and financial data, in particular by making use of new information sources and analytical tools. Second, as users of data to conduct their policies, central banks faced new information needs; this reflected in particular the acute socioeconomic challenges brought about by the pandemic and the unprecedented policy measures taken in response and this needed to be carefully calibrated and monitored.

Following the pandemic, the global health situation appears to have stabilised. One view is that this would lead to some normalisation in the official statistical landscape, reverting to the pre-crisis situation. Another view, however, is that the pandemic has shed light on a number of important information challenges. From this

perspective, it has triggered a fundamental reassessment of the way central banks produce and consume data. The post-pandemic landscape for official statistics could thus be seen as a "new normal", with significant distinctive features vis-à-vis the situation prevailing before 2020.

To shed light on these issues, the 11th conference of the Irving Fisher Committee on Central Bank Statistics (IFC) of the Bank for International Settlements (BIS) held in 2022 was devoted to the "Post-pandemic landscape for central bank statistics". This event was an opportunity to take stock of the new perspectives facing central bank statisticians, especially as regards data production, users' evolving information needs and ways to address them, as well as the challenges that remain.

More than 40 contributions were presented at the event, which was attended by almost 170 participants from 60 countries. These contributions, as referred to in this overview and included in this IFC Bulletin, highlighted the need for central banks to derive adequate long-term lessons from the pandemic, drawing from their dual perspective as producers of official statistics (Section 2) and as key users of economic and financial information (Section 3). They also highlighted the need to deal with a number of long-standing challenges, in particular the impact of globalisation, financial innovation and sustainability issues (Section 4). Fortunately, a number of ongoing projects in the central banking community are addressing these various points, leveraging technical innovation (Section 5) and the potential provided by the wealth of micro-level data available (Section 6). This calls in particular for further progress on the global statistical infrastructure (Section 7) and on liaising effectively with the various stakeholders involved (Section 8).

2. Longer-term impact of Covid-19: the producers' perspective

Covid-19 had immediate, disruptive effects on the statistical production chain, leaving policymakers with acute information gaps in the short run. But the pandemic also underlined the longer-term importance for producers of official statistics to provide more and more varied types of data to their users, not least by leveraging the wealth of information sources available and big data-related tools. These elements can also help to address emerging issues that are still uncovered by existing official statistical frameworks, especially through the production of new types of statistics (Rosolia et al (2021)).

Making sense of the wealth of "organic" data available

Large and sudden disruptions at the height of the pandemic created several difficulties for the producers of official statistics, who were confronted with unavailable sources of information, methodological issues and data quality shortcomings. These challenges were first addressed in a pragmatic way to try to address the most pressing information gaps. But they were also a "wake-up call" for a general review of official statistical frameworks, not least to make a better use of existing data and sources and also of technological innovation in the longer run (Biancotti et al (2021)).

For instance, a number of new big data sources were explored by the Bank of Italy in the course of 2020 to replace missing statistics; this also proved an opportunity

to review the benefits of incorporating new data on web connectivity (eg mobile data), electronic payments (eg credit card transactions) and internet search data (eg Google Trends) to enhance the compilation of the "travel" item of the Balance of Payments (BoP). In particular, mobile phone data can be used to improve estimations on international travellers, as compared with more traditional data sources. Likewise, electronic payment records could be rather useful for data validation purposes, especially to deal with issues related to missing transactions, misclassification and insufficient granularity. Turning to Google Trends indicators, and despite the need to deal with potential breaks in time series when circumstances changes, they can help improve the quality of statistics available on a complementary basis.

More fundamentally, there has been a widespread recognition among central bank statisticians that a wealth of data exist that had been too long neglected . Cases in point relate first to administrative data registers, ie those records created in an organic way as a by-product of government operations. Moreover, financial big data sources have multiplied in recent decades, reflecting the impact of digitalisation in today's economies and the vast amount of commercial and financial operations undertaken every day. Lastly, large data collection efforts have been launched by national authorities and international organisations as part of the lessons drawn from the 2007–09 Great Financial Crisis (GFC). In particular, the Data Gaps Initiative (DGI) endorsed by the G20, and which has covered three phases so far, has been instrumental to closing the most pressing data gaps in financial markets (FSB and IMF (2009, 2015); IMF Staff et al (2023)).

As a result, central banks are working actively on making a better use of the wealth of information available, which is for instance a clear priority in Portugal. This means, first, that they are actively exploring non-conventional, still untapped sources of information. Second, they are leveraging technological innovation to improve the quality of official statistics being produced. In particular, the new statistical production processes can benefit from artificial intelligence (AI) techniques to enhance the accuracy, completeness and consistency of the reporting agents' responses. One example relates to the treatment of regulatory requirements for banks' non-performing loans, for which the Bank of Italy has developed a data quality management (DQM) rule based on past observations (eg number of outliers) to identify quality issues at the observation level automatically and promptly.

Providing more timely and high-frequency indicators

The Covid-19 pandemic also underlined the need for having more high-frequency and timely statistics, not least to assess today's complex economic relationships, monitor evolving fragilities and keep policymakers informed at times of rapid changes and/or very large shocks. In general, traditional official statistics face important delays: to ensure accuracy and comparability, they are compiled on the basis of internationally accepted methodologies and standards and best practices, hence requiring relatively lengthy processes and verifications. In contrast, new types of "alternative" data can be compiled very quickly, not least by relying on innovative techniques, but often at a price in terms of the imperfect quality of the indicators produced.

While this accuracy/timeliness trade-off is not new for public statisticians, the pandemic has clearly changed views on how best to address it. Confronted with "statistical darkness", central banks around the world have recognised the primacy of

providing timely indicators, for instance by mobilising alternative high-frequency data sources, constructing weekly or even daily indicators, and enhancing their nowcasting exercises (IFC (2021a)). From a longer-term perspective, they have also started a more fundamental review of their statistical function with a focus on data production chains. For example, the Central Bank of Türkiye has worked to increase the frequency of private institutions' reporting, from quarterly to weekly, by combining balance sheet data with administrative records. Timeliness has also been improved by simplifying reporting formats. These enhancements have supported the creation of an in-house real time data platform as a monitoring tool of financial market conditions based on well-established processes for data streaming, integration, analysis and visualisation.

An important post-pandemic focus point has been to produce daily data. The value of such indicators is critical for both supporting public authorities' monitoring tasks and providing faster information to society. For instance, the European Central Bank (ECB) has started to compile and disseminate daily data on the overnight interest rate for the euro (the €STR), drawing on the granular information already available in financial markets. This initiative illustrates the potential value of pre-existing statistical micro data in supporting market surveillance and monetary policy as well as in providing reference benchmarks for market participants. Another initiative by the Bank of Israel has been the setup of "new rapid" data sets by complementing various publicly available high-frequency indicators (eg developments in financial and forex markets, electricity consumption) with those available only for internal use (eg banking loans and mortgages in moratorium). The aim was also to make more data accessible to the wider public through interactive dashboards.

Producing new types of indicator

The Covid-19 pandemic also boosted the production of different types of statistics, which can be useful to address new economic and financial issues in the longer run. Central banks have been at the forefront of these initiatives, leveraging on alternative data sources and technical innovation. An obvious example relates to economic agents' opinions, which are increasingly used for policy analysis. Significant efforts have been put on producing indicators on economic sentiment/confidence, based on text-mining tools to summarise qualitative information on agents' perceptions or expectations in real time. While such sentiment indicators were traditionally computed from dedicated surveys of households or firms, the recent approaches have proved more flexible, easier to implement, and better suited for unexpected events – such as during the pandemic when surveys could not be conducted because of lockdowns.

For instance, the Central Bank of Chile has developed a high-frequency sentiment indicator through the application of textual analytics to capture the emotional tone of economic and opinion news, based on the computerised reading of printed media. The approach required the setup of an adequate dictionary to identify different types of emotion, the classification of the various textual inputs, and the benchmarking of the results vis-à-vis other, more traditional types of survey-based confidence indicator. The results have proved particularly useful in assessing the current state of the business cycle, the future evolution of economic conditions, and the degree of confidence among economic agents.

3. Longer-term impact of Covid-19: the users' perspective

A key requirement for official data is to meet the test of practical utility, in the sense that they have to address the information needs of the government, the economy and the public at large.[2] Hence an essential focus for central banks' statisticians in the longer term is to remain relevant to users, including internal decision-makers in their own institutions.

Keeping central bank statistics relevant in the official statistical offering

Beside its short-term impact, the Covid-19 pandemic highlighted that the statistical landscape has to continuously evolve to reflect the changing economic and financial environment. This means that, to be fit for purpose, the official statistics offering needs to remain relevant in view of changing user needs. Among the many related challenges, statisticians need to be able to modify their output rapidly and flexibly; the statistical infrastructure needs to be adaptable and sufficiently granular to meet evolving requests; the trade-off between the costs of producing additional data and their benefits should be carefully assessed; and the purpose of compiling new statistics as a public good should be clearly understood by the public, which calls for strong data governance frameworks.

As regards the contribution of central banks more specifically, the pandemic showed that their data collected on the real economy and the financial system continue to be crucial to fostering monetary and financial stability and serving a wide range of users – eg policymakers, citizens, financial market participants, students and professors, as well as journalists. In that sense, the statistical responsibility of central banks has not changed fundamentally and continues to rely on a number of key elements. First, their data collection exercises must be clearly linked to their mandates. Second, the principles governing official statistics must be rigorously adhered to, not least to ensure that the correct information is used for policymaking. Third, and relatedly, is the recognition that high-quality data are key to supporting the credibility of policy decisions and in turn their effectiveness. Fourth, there must be a focus on securing public trust in this process, although this can be a challenge in the new era characterised by increasingly complex and diverse data sources.

Central bank experience shows that a number of initiatives are being pursued to further consolidate the strategic contribution that their statisticians can have to address users' needs. One important focus area has been the broadening of the toolkits available to manage information, in particular by making the most of the opportunities provided by new big data analytics such as machine learning (ML) to support policy use cases (IFC (2022a)). Another is to strengthen their communication to various user groups, ideally based on both a multichannel and a segmented approach to reach out to the main stakeholders effectively (IFC (2023a)). Third, statistical methodological standards and products need to be constantly updated, for instance to capture more adequately the impact of digital innovation in finance, the issues posed by globalisation, and sustainable development aspects.

[2] See UN (2013), especially its first principle: "(…) official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information".

Addressing internal policy users' needs

Central banks are themselves key consumers of data in conducting their policies. This was particularly obvious during the pandemic, when a wide range of policy measures were taken to mitigate the adverse impact of the crisis on the economy and society. Internal policymakers needed the right information at the right time to properly design, calibrate and monitor their actions (Tissot and De Beer (2021)). One example was Türkiye, where the initial response to the pandemic comprised interest rate cuts, liquidity provisioning, credit facilities and foreign currency swaps to alleviate the financial situation of households and firms. Another example was related to the various monetary and macroprudential measures taken to limit the impact of non-performing loans on the banking system so as to safeguard financial stability at the height of the pandemic, as was the case in Brazil and the United Kingdom.

One key consideration from this perspective was to be able to assess the impact of policy measures in a granular way, ie depending on the groups of economic agents of interest. For instance, the Board of Governors of the Federal Reserve System in the United States has developed "distributional financial accounts" that helped in the monitoring of the situation of different household groups through the pandemic. This analysis showed that, while reduced consumption and large stimulus packages had led to a significant increase in savings in general, important wealth effects were generated by asset price developments, resulting in significant dispersion across households.

Another important issue for central banks as policy institutions was to measure the credibility of the unprecedented measures taken during Covid-19. To address this point, one project at Bank Indonesia led to the construction of a macroprudential policy credibility index, drawing on ML and text-mining techniques based on media news. The aim was to assess the contribution of various factors (eg formulation of the measures, communication) in supporting the credibility of the actions taken to support financial stability. A similar approach was followed to assess authorities' actions in Indonesian payment systems, representing another essential policy area.

Looking forward, central banks' experience suggests that a number of critical aspects need to be considered to further enhance the contribution of public statisticians to evidenced-based policymaking. One is to secure the involvement of subject-matter experts in policy discussions. Second, careful prioritisation is essential, not least in view of resources constraints; this means that the statistical agenda should remain focused on economic and financial issues, the bread and butter of central banking. Third, international collaboration is key, primarily among central banks especially in the context of the IFC but also with other stakeholders eg international organisations, National Statistical Offices (NSOs), statistical associations such as the International Statistical Institute (ISI), and academia. Such cooperation can be instrumental to support the sharing of knowledge and experience as well as the global application of internationally recognised statistical standards. And, lastly, these efforts should be accompanied by a better use of the opportunities offered by micro data (Israël and Tissot (2021)).

4. Facing (not so new) challenges

In addition to addressing the longer-term lessons of the pandemic as regards the supply and demand of data, recent years have also highlighted the urgent need for

official statisticians to deal with a number of long-standing challenges. This calls in particular for central bank statistical frameworks to be adapted to better reflect structural developments in the global economy, especially as regards globalisation, digitalisation in finance, and sustainability aspects such as climate change.

Globalisation

An important challenge has been how to address the impact of globalisation on the official statistical framework. Obviously, it has become increasingly difficult to measure the footprint of multinational enterprises (MNEs) in domestic economies, due to the size of cross-border operations related to international supply chains, the complexity of corporate global ownership structures and the opaqueness of their restructuring strategies not least for tax optimisation purposes (Francois and Vicard (2023)). These elements have clearly complicated the work of official statisticians, especially in central banks, which are typically the main compilers of external sector statistics and are responsible for maintaining direct investment registers in many countries (IFC (2020a)).

Experience shows that the main difficulties faced in the compilation of external sector statistics often relate to MNEs' four main types of data deficiency. First, the limited availability and timeliness of group-level sources data, since disclosed financial statements appear insufficient to support statistical compilation and need to be refined and validated with complementary data sets. Second, information on cross-border operations often varies between national compilers, reflecting multiple factors such as limited data-sharing and the lack of comprehensive and internationally agreed guidelines (eg for the treatment of mergers). This can lead to important inconsistencies between national BoP figures, leading to sizeable quality issues in external statistics (eg size of the errors and omissions item). Third, the treatment of special corporate restructuring cases has become increasingly complex, due for instance to the difficulties of identifying the location of corporate group activities within national borders and the role played by ad hoc units in global financial centres (eg Luxembourg), sometimes only on a temporary basis. Lastly, national data collection systems are relatively rigid, hampering statisticians' ability to track MNEs' evolving operations flexibly and quickly.

These limitations are clearly obvious for the compilers of foreign direct investment (FDI) statistics, especially because of the difficulty of correctly identifying MNEs' chains of control. Complex corporate structures, with layers of intermediate holding companies, imply that the identity of the final investors is often masked. Reflecting these limitations, FDI data sets are usually compiled on the basis of the Immediate Investor Country (IIC) principle rather than the Ultimate Investing Country (UIC) one. This represents an important data gap, since the correct identification of the residence of the entities ultimately controlling FDI can be essential to understand intragroup economic relationships and improve the traceability of cross-border funding. It also complicates the interpretation of published FDI data, not least because of the inflated role played by intermediate financial centres in the global economy (Pogliani et al (2022)). Lastly, the above difficulties posed by globalisation can also have a huge impact on aggregated national accounts statistics (eg GDP, investment) especially in small open economies, as seen with the 2015 "Irish case" (OECD (2016)).

Digitalisation and financial intermediation

Another long-lasting challenge highlighted by the pandemic, in particular during the market turmoil observed in early 2020 (FSB (2020)), is the impact of innovation on financial intermediation and financial auxiliary functions, especially through the operation of non-bank entities. The so-called "fintech" topic, initially defined by the Financial Stability Board (FSB) as technologically enabled financial innovation that could result in new business models, applications, processes or products with an associated material effect on the provision of financial services (Carney (2017)), has indeed been on the agenda of the central banking community for many years. In particular, the fast-growing scope of fintech activities requires their adequate classification in statistical systems, not least to allow for their correct monitoring by public authorities. But a key obstacle is that there is currently no unique, internationally harmonised definition of fintech for statistical or other classification purposes. It is thus essential to adapt statistical frameworks to better address the challenges posed by digital innovation in finance through the correct identification of the related fintech providers and products (IFC (2020b)). International organisations have been actively working on these issues, not least in the context of the DGI as well as of the currently ongoing updating of international statistical standards – eg the System of National Accounts (SNA), the Balance of Payments Manual (BPM) and the International Standard Industrial Classification of All Economic Activities (ISIC); see Baer et al (2022).

Yet, three main issues stand out in this context. One is that a number of fintech firms (eg IT services providers) can be located outside the sector of financial firms, calling for more detailed classification in statistical systems and/or enhanced methodological guidance to solve the related identification problems. Another issue relates to the appropriate, feasible and consistent treatment in macroeconomic financial accounts of the digital assets used as a means of payment – eg central bank digital currencies (CBDCs), stablecoins and other types of cryptoassets. Lastly, an important concern for central banks, that are tasked in many jurisdictions with the mandate of promoting financial inclusion and supervising payment systems, is to monitor the use of new financial services by the population and the correct understanding of the risks involved.

Sustainable development

The pandemic has clearly reinforced the need to pay greater attention to sustainable development, with a particular focus on its environmental, social and governance (ESG) aspects. This calls for more internationally comparable statistics, based on clearly established standards and definitions to reliably inform investors, for instance to support the transition to a less carbon-intensive economy. But the current lack of international standardisation of ESG statistical frameworks and taxonomies can reduce the expected benefits of upcoming disclosure requirements, hampering the gathering of comprehensive information and the proper alignment of market incentives.

As regards central banks and supervisory authorities more specifically, the most pressing data needs appear to relate to their activities in the areas of financial stability, asset and reserve management and monetary policy, reflecting the importance of ESG issues for the solvency of financial institutions (IFC (2022b)). They have thus been at the forefront of efforts to develop sustainable finance statistics, focusing mainly on

establishing analytical frameworks, designing sustainability indicators and actual monitoring (IFC (2021b)). In particular, the work of the Network on Greening the Financial System (NGFS) of central banks and financial supervisors has been instrumental to identify relevant information sources and address acute data needs (NGFS (2021)).

Despite these various efforts, important shortcomings remain to be addressed, especially as regards the availability and quality of granular-level ESG information. One view is that more than 100 indicators may be needed to fulfil existing information gaps. But, as emphasised by a recent study conducted by the Bank of Spain, available ESG micro-level data suffer from several limitations such as limited coverage of individual firms, heterogeneity in the standards and definition of indicators, lack of digitalisation, and verification and regulatory limitations. As a result, there is an obvious lack of granular firm/asset-level data to meaningfully measure the carbon footprint of economic and financial activities. Another data gap refers to forward-looking climate risk indicators (eg emission pathways), noting that most of the currently available ESG indicators are backward-looking. This is an important challenge, since forward-looking indicators would be key to facilitate tracking commitments towards a greener economy.[3]

5. Looking forward: leveraging innovation…

The above developments suggest that central bank statisticians may be facing a quite heavy agenda should they want to both address the new lessons of the pandemic as well as longer-term challenges. Fortunately, a number of factors will support their efforts.

An obvious one is innovation. The new normal for central banks statisticians is likely to rely heavily on the use of data science to perform their traditional tasks (Nymand-Andersen (2021)). The pandemic has clearly boosted interest in using "big data", ie the multiple types of information sets often characterised by a high degree of volume, variety and velocity – the so-called "Vs" (Laney (2001)), although the big data sources of interest to central banks can be quite varied in practice, ranging from large traditional micro data sets to internet-based indicators and unstructured information such as text (IFC (2017)). Moreover, innovative analytical techniques, such as ML and other AI-based tools, can improve the efficiency and effectiveness of statistical work, by facilitating the gathering of more and better information and the dealing with complex and new complementary sources in a more automated way (IFC (2022a)). This can bring multiple benefits, in terms of granularity, flexibility, timeliness, accuracy and efficiency. For instance, it can support central bank statisticians in their quest to capture data in real time from existing sources, and help them construct innovative indicators to cover new areas of interest – see the recent project developed in Indonesia to gauge users' satisfaction with applications for mobile payments services, which can effectively support financial inclusion.

[3] See NGFS (2021): "Given the importance of forward-looking assessments of both physical and transition risks, the current reliance on mostly backward-looking data is unsatisfactory". There is thus a "need to understand the point-in-time performance of an exposure against a transition pathway – hence the need for firms to disclose their transition plans – as well as the impact of adaptation and mitigation measures on the evolution of the risks."

A first and immediate benefit is to strengthen the core statistical function of central banks. This can facilitate the dealing with the larger and more complex data sets that are in increasing demand to address the needs for more granular, timely and high-frequency information – not least to deal with events such as Covid-19. Moreover, several projects have highlighted the usefulness of ML tools for supporting DQM purposes in comparison with more traditional modelling approaches (Chakraborty et al (2017); Maddaloni et al (2022)). Better data quality can be instrumental to securing trust in the entire statistical production chain, from data collection to dissemination, and to strengthening central banks' reputations not only as data producers but also as policymakers taking action based on the right information.

Indeed, the contribution of more and better data and advanced tools can effectively support central banks' actions in multiple areas, such as macroeconomic analysis and forecasting, financial market monitoring and financial risk assessment (IFC (2019a)). As regards, for instance, the use of big data analytics to forecast macroeconomic indicators in near real time, a recent project pursued at the Bank of Italy aimed to use granular administrative data on motor vehicle registrations to "nowcast" business investment in transport equipment and hence capital goods. The approach also allowed for a distinctive approach across regions, which proved useful in tracking the impact of the pandemic. Another example is the construction by Bank Indonesia of a proxy for household consumption from retail payment system transaction data, based on text-mining techniques. This advanced indicator could be compiled within a few days after the end of the reference period and showed a high correlation with published (with significant lags) official household consumption numbers.

Yet the use of big data sources and tools is not without challenges. The new data sets may provide a false sense of accuracy, as they can present substantial hard-to-assess selection biases that limit their representativity of the whole population of interest (Mehrhoff (2019)). Similarly, AI tools trained on preclassified and/or inaccurate data sets may be of little help and raise the risk of "algorithm unfairness" (IFC (2021c)). Another risk is complexity, compounding the general literacy challenges faced by users of official statistics. This calls for the avoidance of "black box" messages and the development of innovative tools to inform policymakers and also researchers and the broader public. A final challenge is privacy protection and the respect of confidentiality settings, another key principle in official statistics.[4] Respecting this principle can be challenging in practice when dealing with large data volumes with multiple variables, not least because of the risk of re-identification of anonymised data posed by the increased IT computing power available and the multiplication of cross-sectional and longitudinal databases that could be interlinked (IFC (2023b)).

Addressing the above issues puts a premium on adopting strong and comprehensive data governance frameworks (IFC (2021d)). It also calls for a clear recognition of the various trade-offs involved, not least given the resource constraints (eg budget, skilled staff, IT infrastructure) faced by public institutions. This means striking the right balance between costs and benefits when compiling data from new sources (for instance in term of the timeliness/accuracy trade-off), as compared with

[4] See UN (2013), especially its sixth principle: "Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes."

making better use of existing data. Lastly, users need to have a clear and comprehensive picture of the analytics provided, including the degree of uncertainty associated with the data and techniques used. Any shortcomings of the new big data and tools used should thus be communicated transparently to them, not least to highlight the degree of uncertainty surrounding the environment and hence policy decisions.

6. … making a better use of micro data…

A clear message from the central banking community is that many of the challenges highlighted by the pandemic could be addressed by making better use of the large amount of micro-level information already available in today's societies (IFC (2021e)). The growing supply of granular data reflects the impact of digitalisation (the digital footprints left by a large number of activities), the importance of government activities (eg the development of administrative data sets), and the statistical follow-up to the last GFC as well as the Covid-19 pandemic – with the launch of a number of coordinated initiatives such as the DGI to collect more granular economic and financial data to support decision-makers. Supply was also spurred by technology innovation, which has clearly facilitated the collecting, processing, validation, integration and analysis of large data sets. This has further boosted the use of existing microdata and the compilation of new statistical indicators as well as the application of innovative analytics. And, indeed, various central bank projects have highlighted the benefits of making a better use of the wealth of micro-level information, in particular to deal with globalisation, ESG issues, and financial stability.

As regards first globalisation, one initiative developed in the context of the European Committee of Central Balance Sheet Data Offices (ECCBSO) has highlighted the new possibilities offered by the integration of micro-level indicators on firms' performance and group structures to generate nationality-based statistics (ie based on the country of control principle) that can clearly facilitate the assessment of the footprint of global groups outside their jurisdictions of residency (Tissot (2016)). Another focus has been on analysing the international chain relationship for FDI, which is often intermediated by financial centres. For instance, the Bank of Romania has used granular data from FDI registers to clarify the global network of direct investment linkages across different types of firm by industry and the impact of investment decisions made by their ultimate controlling parents. Similarly, an ECB study was conducted to assess who stands behind European FDI investors using company-level data on corporate ownership structure to determine the importance of non-EU corporations in ultimately controlling FDI activity, including its intra-EU component. Turning to the Bank of Italy, it has developed a methodology based on innovative techniques to identify ultimate hosting economies in bilateral FDI statistics, which is often a challenge due to lack of detailed data. The ECB has also developed a specific metric (the "significance multiplier") to enhance the quality of the information on group structures contained in its granular Register of Institutions and Affiliates Data (RIAD). The approach has also helped to evaluate changes in corporate structures motivated by specific business decisions (eg mergers) on a high-frequency basis. Lastly, recent geopolitical uncertainties have raised questions about the pace and/or features of globalisation and highlighted the usefulness of micro data to better understand these developments. One question relates for instance to the possibility of "near-shoring"/"friend-shoring", ie the reconfiguration of global value chains in favour of countries that are geographically closer (Panetta (2023)). This kind

of analysis may be difficult to pursue based on macro data solely, and one solution can be to use non-traditional data sources instead, such as expatriate travel, land prices in key areas or transportation volumes.

Turning to the ESG area, a number of granular data sets exist that can shed useful light, as highlighted by the Bank of Spain's project to estimate the carbon footprint of banks' credit portfolios and track the evolving impact of their loans. Moreover, a key objective has been to properly analyse the consequences of sustainability issues on the financial system, with two types of need: first, to assess the impact of ESG factors on the risk profile and performance of specific institutions; second, to measure the contribution of sustainable finance – understood as the process of taking ESG considerations into account when making financing decisions with the aim of fostering long-term investment in sustainable economic activities (IFC (2022b)). Indeed, as highlighted by a recent review by the Central Bank of the United Arab Emirates (UAE), a growing number of institutions already integrate ESG criteria into their investment strategies. The use of green finance instruments is in fact likely to increase further, not least in view of the development of the Principles for Responsible Investment by an UN-supported international group of institutional investors. Similarly, the Deutsche Bundesbank has analysed the investment strategy of self-proclaimed ESG exchange-traded funds (ETFs) using company-level data to analyse the impact of their financing; in this context, accessing granular information proved particularly important to cope with the diversity of sustainability objectives pursued by the various types of fund. A parallel approach allowed for the extraction of relevant forward-looking risk indicators at a sectoral and country level, using existing micro data available from a private provider that helped to overcome the challenges posed by, in particular, limited comparability across data sources and over time.

A third key contribution of micro data to support central bank policies is to shed light on the functioning of the financial markets especially as regards idiosyncratic fragilities and contagion effects. For instance, the UAE Central Bank has developed a holistic risk-based approach to assess banks' real estate sector exposures and concentration risks. Similarly, the Bank of Israel has worked to identify potential vulnerabilities based on a granular data set of assets' holdings of institutional investors, which allowed the analysis of diversification effects, systemic factors, and the influence of initial conditions. More generally, granular reports from supervised institutions can facilitate the monitoring of specific markets that play an important role in financial stability. For instance, the Bank of Japan has utilised granular transaction data to analyse the market structure and haircuts for repo and securities lending markets. This type of information is highly valuable for authorities willing to take adequate measures preventively and/or in reaction to specific events. It also puts a premium on the automated detection of reporting errors, as done for instance at the Bank of Italy for insurance companies' reports through ML techniques[5] – underscoring again the value of advanced analytical tools to support micro-DQM processes efficiently, in particular when human resources are limited.

[5] The related paper from the Bank of Italy on "Statistical matching for anomaly detection in insurance assets granular reporting" (La Serra and Svezia (2022)) received the IFC award for the best paper presented at the conference by a young statistician.

7… making progress in statistical infrastructure…

The central bank community is supporting initiatives to improve general aspects related to the global statistical infrastructure (eg registers, identifiers and statistical standards). One example relates to the Legal Entity Identifier (LEI), a 20-digit reference code to uniquely identify legally distinct entities that engage in financial transactions, which has already become mandatory in several jurisdictions. It has been argued that this standardised and unique entity identification played an important role during the Covid-19 crisis, for instance in the United Kingdom by addressing (i) the difficulty of verification of merchants on online platforms and (ii) the drain on public sources due to fraudulent applications for public funds that were designed to support firms (GLEIF (2020)).

An important focus has been on the methodologies supporting official statistics compilation, since central banks' data production has to take place in the general context of globally standardised methodologies. This requires continued international collaboration to support the harmonisation of standards and the exchange of learning experiences, not least within the central bank community but also with relevant stakeholders at national (eg NSOs) and international levels (eg OECD, IMF and the UN). A key goal of the IFC is thus to promote knowledge-sharing and international cooperation on statistics-related methodologies, initiatives and training (see IFC (2023c)). Another important dimension is to further develop statistical standards and support related ongoing initiatives – especially in the context of the new DGI and of the revision of international statistical manuals. This calls also for improvements in the usability of these manuals, whose lifecycles often run over several decades, for instance by developing digital versions that can facilitate navigation across different domains and improve methodological consistency (eg reduction of overlaps and inconsistencies). This would also allow for flexibly adding methodological extensions as new, specific statistical issues arise.

One particular topic of interest to central banks in this context relates to globalisation, with the current updating of the IMF Balance of Payments and International Investment Position manual (BPM6), which aims in particular to improve the statistical treatment of corporate restructuring operations in external statistics and to favour the reduction of bilateral asymmetries. Another key domain relates to the fintech universe, which combines novel types of financial service with more traditional ones performed in novel ways and activities that are not financial. This calls for a proper classification of the various types of activity involved (eg cryptoassets, banktech, insurtech, tech facilitators/infrastructure providers).

Some challenges may also require a more fundamental overhaul of conceptual approaches. A case in point relates to the residency principle (country of location) underlying national statistics (IAG (2015)). Since MNEs' operations cannot be captured comprehensively by isolated, country-based statistical systems, nationality-based statistics could be developed as a useful complement. This approach has been already followed for the compilation of a number of BIS statistical products, such as the international debt security statistics (IDS) and the international banking statistics (IBS). But it has to be reinforced by a common statistical infrastructure supporting the cross-border sharing of firm-level information, as already organised in the context of financial institutions under the BIS umbrella (Tissot (2017)). Fortunately, similar initiatives are also under way as regards information on non-financial corporates, especially by the OECD for MNEs activities (AMNE Database) and Trade in Value Added (TiVA Database) as well as by Eurostat for statistics on the types of enterprise

engaged in international services – the Services Trade by Enterprise Characteristics (STEC) experimental statistics, produced by linking international trade in services micro-data with the business register at enterprise level.

8… and fostering cooperation

The complex and various challenges facing official statisticians suggest that central banks cannot act in isolation and should continue to pursue their cooperation with relevant stakeholders.

Reaching out to all relevant stakeholders

A key group of interest to central banks statisticians are their counterparts in national statistical systems (NSS), especially NSOs, as well as the statistical departments of international organisations. Their unique expertise is instrumental for developing data standards and new methodologies. Moreover, good cooperation helps reduce overlaps and maximise synergies, an important issue for public authorities facing resource constraints. This collaboration can take various forms, from the exchange of experience and technical assistance to the setup of digital platforms for information-sharing, including statistical repositories of models (where to find data and metadata). For instance, the NGFS has emphasised the importance of setting up a dedicated repository of climate data to support financial sector stakeholders to share best practices in identifying related data need, sources and gaps. Similarly, the IFC has actively promoted the exchange of experience in the official statistical community, building on its broad, global membership of central banks and its position as an affiliated member of the ISI.

Another important group of stakeholders for central banks relates to the private data providers that are playing an increasingly role in the current search for alternative data sets, especially for the purpose of closing existing gaps in official statistics. One essential area relates to ESG issues, for which a good understanding of both the information market and the interests of alternative data providers is essential. Establishing partnerships can be the best way to access these data in a continuous way, design related regulations and standards to prevent monopolies, and properly incentivise the private sector to develop, create and deliver novel (technical) solutions for the provision of new indicators. For their part, public statisticians have to improve the useability of these complementary data so that they can be incorporated into the official statistical framework in an agile and timely manner. This can be achieved, for instance, by embedding public statisticians in external data providers so that they can better understand the management and supply of their alternative indicators.

A third important target group are the users of statistical information. Central banks, like other compilers of official statistics, are well aware that communication is a core statistical task to ensure that the data are fit for purpose and that their value is maximised. This puts a premium on developing statistical literacy in the population, ensuring that users understand and accept what statisticians are doing, and securing public trust in the overall official statistical framework.

Facilitating data access for research purposes

As a key user group targeted by central banks, academic researchers study internal granular data sets for scientific and analytical purposes – a demand that has been growing rapidly in recent years. However, accessing such detailed information can be quite sensitive because of the confidentiality and privacy issues involved.[6] The rules and practices for protecting confidentiality have thus to be revisited to make micro data more accessible. Two key workstreams are involved here: first, the clarification of the procedures defining how authorised personnel can use the data; this aspect has become a key element of central banks' governance frameworks. Second, the development of specific processes for data dissemination, including research code-sharing, remote execution of computation exercises based on confidential data, and in-house dedicated data research centres.

To shed light on these issues, several central banks with a number of NSOs and international organisations have been involved in INEXDA, the International Network for Exchanging Experience on Statistical Handling of Granular Data (IFC (2019b)). Its work comprises a stocktaking of the granular data sets available across countries, the review of best practices for granting access to open software solutions and data and metadata, and the identification of common features across jurisdictions with a view to the potential harmonisation of data access procedures. A key focus has been on the exchange of experience on the accessibility of micro administrative data and related data protection techniques, with the development of machine-readable description of access procedures, the review of the technical, administrative and organisational features for accessing granular data for research purposes, and the promotion of a metadata schema to provide metadata for microdata on the data set level (Bender et al (2019)). This schema is based on (i) the data set aspects of access procedures (ie information collected to unambiguously identify a data set, its origins, and its access modes); (ii) information regarding the aggregation of granular data, especially when it involves combining different data sets; and (iii) information on users of this information and related projects.

The INEXDA initiative has highlighted a number of important features related to the governance of data access for research purposes. First, output control has typically become mandatory for non-publicly available data, and is usually performed by researchers and subsequently checked by the data providers. Second, the often preferred access mode is on-site access, when researchers have to work in a secure environment based in the premises of the institution providing the data. Third, access modes can differ depending on the available data sets or the degree of anonymisation required: in particular, while formal anonymisation appears to be the most commonly used method, there can be various approaches governing data access for scientific research, with the involvement of different decision bodies in these processes (supported by specific confidentiality agreements or memorandums of understanding) and based on various techniques – eg remote computation,

[6] One typically distinguishes between the risks of "direct disclosure" (ie disclosure of information with identifiers) and "statistical disclosure" (ie the risk of re-identification of anonymised data, a risk that is growing with the increased IT computing power available and the multiplication of cross-sectional and longitudinal databases that could be interlinked); see National Research Council (2005).

identity masking tools, use of so-called public use files (PUFs) that consist of micro-level records prepared in such a way that individual entities cannot be identified.[7]

Such knowledge-sharing exercises have highlighted the key role of data protection procedures to protect confidentiality (UNECE (2019)). One example is the Statistical Disclosure Control (SDC) method developed at the Bank of Spain's BELab data laboratory to ensure the anonymisation of sensitive time series microdata such as loan-level information. The approach is based on an extensive use of big data analytics and aims to mitigate the risk of disclosing the identity of specific borrowers resulting from the combination of multivariate information, since debtors may have multiple loans and loans may have multiple debtors (hence raising the risk of debtor re-identification even when the data set has been anonymised).

Another important area relates to code-sharing, which is increasingly considered a best practice in empirical scientific research, not least for transparency and accountability reasons. In particular, a growing number of journals employ data editors to re-run code and validate results of published articles. The implication is that authors have to provide all the data, code, and processing instructions necessary to exactly replicate the analysis and obtain identical results. But, as highlighted by a recent pilot project undertaken for the UK NSO Secure Research Service, the wider sharing of code outside a closed researcher environment requires the safeguarding of statistical confidentiality, the validation of its quality, and adequate disclaimers to protect the reputation of data owners' and publishers. Moreover, the reproducibility of research results obtained in a trusted environment is not without challenges, as it requires the review of the related data access rights, the consideration of supporting auxiliary material that has not been disclosed, and the review of the code to be shared outside the secure environment.

A final lesson is that preparing research-ready data can require a significant level of resources. These are required to deal with large (numbers of) data sets, complex data structure (eg multiple identities), and heterogeneous legal frameworks (eg when specific rules depend on the data considered). Hence, the onboarding of external researchers in internal research data centres (RDCs) can be cumbersome and resource-intensive. This puts a premium on developing tools to support the pre-selection of the information that can be shared in a more automatised way, an approach currently under consideration at the Deutsche Bundesbank.

[7] Noting that there are many ways in practice to protect data confidentiality (IFC (2023b)).

References

Baer, A, F Berry and W Phiri (2022): "Update of statistical standards and implications on the calibration of statistical business registers", *IMF Statistics*, 29 September.

Bender, S, B Hausstein and C Hirsch (2019): "An introduction to INEXDA's metadata schema", *IFC Bulletin*, no 49, January.

Biancotti, C, A Rosolia, G Veronese, R Kirchner and F Mouriaux (2021): "Covid-19 and official statistics: a wakeup call?", *Bank of Italy Occasional Papers*, no 605, April.

Carney, M (2017): "The promise of fintech – something new under the sun?", speech at the Deutsche Bundesbank G20 Conference on Digitising finance, financial inclusion and financial literacy, Wiesbaden, 25 January.

Chakraborty, C and A Joseph (2017): "Machine learning at central banks", Bank of England, *Staff Working Papers*, no 674, September.

Financial Stability Board (FSB) (2020): *Holistic Review of the March Market Turmoil and COVID-19 pandemic: Financial stability impact and policy responses*, November.

FSB and International Monetary Fund (IMF) (2009): *The financial crisis and information gaps*.

——— (2015): *The financial crisis and information gaps – Sixth Implementation Progress Report of the G20 Data Gaps Initiative*.

Francois, M and V Vicard (2023): "The most complex multinationals engage in tax avoidance", *La Lettre du CEPII*, no 434, May.

Global Legal Entity Identifier Foundation (GLEIF) (2020): *Response of the GLEIF to the UK Department Digital, Culture, Media and Support Consultation on the National Data Strategy*, December.

IMF Staff, Inter-Agency Group on Economic and Financial Statistics (IAG) and FSB Secretariat (2023): *G20 Data Gaps Initiative 3 Workplan – People, planet, economy*, March.

Irving Fisher Committee on Central Bank Statistics (IFC) (2017): "Big data", *IFC Bulletin*, no 44, September.

——— (2019a): "The use of big data analytics and artificial intelligence in central banking", *IFC Bulletin*, no 50, May.

——— (2019b): "INEXDA – the Granular Data Network", *IFC Working Papers*, no 18, October.

——— (2020a): "Bridging measurement challenges and analytical needs of external statistics: evolution or revolution?", *IFC Bulletin*, no 52, October.

——— (2020b): "Towards monitoring financial innovation in central bank statistics", *IFC Report*, no 12, July.

——— (2021a): "New developments in central bank statistics around the world", *IFC Bulletin*, no 55, November.

——— (2021b): "Sustainable finance data for central banks", *IFC Report*, no 14, December.

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

——— (2021d): "Issues in data governance", *IFC Bulletin*, no 54, July.

——— (2021e): "Micro data for the macro world", *IFC Bulletin*, no 53, April.

——— (2022a): "Machine learning applications in central banking", *IFC Bulletin*, no 57, November.

——— (2022b): "Progress, challenges and recent innovations in sustainable finance", *IFC Bulletin*, no 56, May.

——— (2023a): "How central banks communicate on official statistics", *IFC Report*, no 15, February.

——— (2023b): "Data-sharing practices", *Guidance Note*, no 3, March.

——— (2023c): *Annual Report*, January.

Inter-Agency Group on Economic and Financial Statistics (IAG) (2015): "Consolidation and corporate groups: an overview of methodological and practical issues", *IAG reference document*, October.

Israël, J-M and B Tissot (2021): "Incorporating microdata into macro policy decision-making", *Journal of Digital Banking*, vol 6 no 3, January.

La Serra, V and E Svezia (2022): "Statistical matching for anomaly detection in insurance assets granular reporting", *IFC Working Papers*, no 22, October.

Laney, D (2001): "3D Data Management: Controlling Data Volume, Velocity and Variety", *META Group Research Note*, 6, February.

Maddaloni, P, D Continanza, A del Monaco, D Figoli, M Di Lucido, F Quarta and G Turturiello (2022): "Stacking machine-learning models for anomaly detection: comparing AnaCredit to other banking datasets", *Bank of Italy Occasional Papers*, no 689, April.

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

National Research Council (2005): *Expanding Access to Research Data: Reconciling Risks and Opportunities*, Washington DC.

Network on Greening the Financial Sector (NGFS) (2021): *Progress report on bridging data gaps*, May.

Nymand-Andersen, P (2021): *Data Science in Economics and Finance for Decision Makers*, Risk Books.

OECD (2016): "Irish GDP up by 26.3% in 2015?", October.

Panetta, P (2023): "United we stand: European integration as a response to global fragmentation", speech at Bruegel event, Brussels, 24 April.

Pogliani, P, G von Peter and P Wooldridge (2022): "The outsize role of cross-border financial centres", *BIS Quarterly Review*, June.

Rosolia, A, S Stapel-Weber and B Tissot (2021): "New developments in central bank statistics around the world", *Statistical Journal of the IAOS*, vol 37, no 4, pp 1055–60.

Statistical Journal of the International Association for Official Statistics (IAOS) (2022): "The involvement of the international statistical organisations in measuring the impact of COVID-19", vol 38, no 3, *Special Issue*, January.

Tissot, B (2016): "Globalisation and financial stability risks: is the residency-based approach of the national accounts old-fashioned?", *BIS Working Papers*, no 587, October.

——— (2017): "Data-sharing: what, why and how? The BIS experience", presentation at the Third Meeting of the Financial Information Forum of Latin American and Caribbean Central Banks, October.

Tissot, B and B de Beer (2021): "Implications of Covid-19 for official statistics: a central banking perspective", *Theoretical Economics Letters*, vol 11 no 4, August.

UNECE (2019): *Principles and guidelines on confidentiality aspects of data integration undertaken for statistical or related research purposes*, Geneva.

United Nations (UN) (2013): "Fundamental Principles of Official Statistics", resolution adopted by the Economic and Social Council, E/RES/2013/21, 28 October.