Are the post-crisis statistical initiatives complete? An overview

Overview of the ninth IFC conference[1]

Evelyn Truong and Bruno Tissot[2]

"*Are post-crisis statistical initiatives completed?*" This was the topic of the ninth conference of the Irving Fisher Committee on Central Bank Statistics (IFC), hosted at the Bank for International Settlements (BIS) on 30–31 August 2018, and attended by almost 150 participants from more than 50 countries. The question is key for today's central bank statisticians. The Great Financial Crisis (GFC) of 2007–09 brought to light the need to address a wide range of data issues, such as the information requirements associated with globalisation, non-bank financial intermediation, complex interbank relationships, and macroprudential policy needs. This has led to **numerous statistical initiatives** (Borio (2013)), a number of them still in their implementation phase a decade after the GFC. Yet, the conference was also an occasion to take stock of more recent developments, especially those related to information-sharing, data dissemination and the impact of technological innovation – such as the digitalisation of the economy and the opportunities and challenges of big data and machine learning in financial statistics.

Statisticians around the world have made considerable progress towards filling known blind spots in data collections since the GFC, as stressed by BIS General Manager Agustín Carstens (BIS) in his opening remarks. Particularly noteworthy have been the efforts of several international organisations, including the BIS,[3] as well as national authorities in the context of the G20-endorsed Data Gaps Initiative (DGI).[4] However, collecting high-quality data may not be sufficient to promote financial stability and prevent the next financial crisis. What is crucial is **the lens that is applied to the data**: connecting the dots is even more important than collecting them (Caruana (2017)). Moreover, official statistics should be constantly evolving to reflect the changing financial environment. This is of particular relevance in the context of fintech's[5] rapid expansion and the shift to financing via bond markets and non-bank

[1] The views expressed here are those of the authors and do not necessarily reflect those of the Bank for International Settlements (BIS), the Reserve Bank of New Zealand (RBNZ) or the Irving Fisher Committee on Central Bank Statistics (IFC).

[2] Respectively Senior Analyst (Economics), RBNZ (Evelyn.Truong@rbnz.govt.nz); and Head of Statistics and Research Support, BIS, and Head of the IFC Secretariat (Bruno.Tissot@bis.org).

[3] Especially in the context of the Inter-Agency Group on Economic and Financial Statistics (IAG), which comprises the BIS, the European Central Bank (ECB), Eurostat, the International Monetary Fund (IMF, Chair), the Organisation for Economic Co-operation and Development (OECD), the United Nations (UN) and the World Bank (WB). The IAG was established in 2008 to coordinate statistical issues and data gaps highlighted by the global crisis and to strengthen data collection.

[4] See FSB-IMF (2009) for the first phase of the DGI; and FSB-IMF (2015) for its second phase (2016–20).

[5] Defined as "technology-enabled innovation in financial services that could result in new business models, applications, processes or products with an associated material effect on the provision of financial services" (FSB (2017)).

financial intermediaries (Shin (2013)). The bottom line is that central bank statisticians should be able to adapt continuously and flexibly to an evolving financial world.

As the global financial environment is changing, so too are policy needs. As emphasised by the IFC Chair, Claudia Buch (Bundesbank), in her keynote address, a lot has happened in the decade since the GFC: data gaps have been closed, new technologies have transformed the production of statistics, and structured evaluations of post-crisis financial sector reforms have started at the G20 level. Yet, in order to sustain these efforts and transform "data" into "information", three points need to be emphasised. First, central bank statistics are of interest to many stakeholders (analysts, researchers, policymakers and private market participants), but they will be used only if they are easily accessible. Communication with stakeholders, a clear legal basis to support data-sharing, the use of joint methodological standards, and measures to protect confidential information are thus crucial (IFC (2015a)). Second, when discussing new data initiatives, the costs and benefits need to be weighed up carefully. Detailed, granular data improve the quality of analytical work, and thus contribute to better policies and public welfare. But reporting of statistical and supervisory information can be costly, particularly if data initiatives are not well coordinated. Improving this trade-off between costs and benefits requires long-term strategic planning and consultation with industry. Third, policy analysis and statistical work can be integrated and coordinated more closely. Before launching policy projects, data needs should be carefully considered, and sufficient time should be budgeted for data work. Statistical initiatives have to take into consideration the type and granularity of data needed to conduct good causal evaluation studies. Both require close coordination between statistical and policy departments in central banks.

The presentations at the conference shed useful light on these various issues. The related papers, spread across nine different sessions[6] and referred to in this overview, are included in this IFC Bulletin. Their findings can be summarised along with **three main messages**. First, significant efforts have already been made to address most of the known post-crisis information gaps, although the related data collection exercises remain unfinished and will take considerable time to be completed. Second, turning data into useful information requires the effective alignment of statistical collections and policy objectives; to this end, one should focus on data-sharing arrangements and techniques for presenting and disseminating information. Third, new data needs have emerged as a consequence of financial digitalisation and technical innovation: these new needs will surely play a major role in shaping central bank statistics looking ahead.

1. Addressing post-crisis data gaps

The past decade of focused and coordinated efforts by the central bank statistical community has substantially improved the scope and quality of financial data

[6] Chaired by, respectively, Hock Chai Toh (Central Bank of Malaysia), Robert Kirchner (Deutsche Bundesbank), Naruki Mori (Bank of Japan), Pedro Duarte Neves (Bank of Portugal), Roh Chung Seak (Bank of Korea), Luis Teles Dias (Bank of Portugal), Joe McNeill (Central Bank of Ireland), Gülbin Şahinbeyoğlu (Central Bank of the Republic of Turkey) and Aurel Schubert (Vienna University of Economics and Business).

available. To give a few examples, the understanding of the global interlinkages between both countries and sectors has improved, more granular data have become available for monitoring the banking sector, and better information is being collected to measure the size and the systemic role of non-bank financial intermediaries.

That said, many statistical initiatives to address the information gaps identified in the wake of the GFC are still under way, and time and resources will be needed to complete the related data collection exercises. The conference highlighted three major areas of interest: the impact of globalisation, rising non-bank financial intermediation, and the need for more granular information.

The impact of globalisation

Many of the critical data gaps identified during the GFC were associated with the impact of globalisation, reflecting the **difficulties in measuring the footprint of global groups** outside their jurisdiction of residency (see Tissot (2016a)). As underscored by Peter van de Ven (OECD) in his keynote address, this is a major obstacle hampering the measurement of domestic activity, for instance when calculating the headline indicator for GDP. This calls for the System of National Accounts framework (SNA; European Commission et al (2009)) to be enriched with a far broader set of statistics. Of course, statisticians' efforts should not be limited to the analysis of the impact of globalisation. There is a more general need to better understand the functioning of the economy, for instance by considering unpaid household activities, integrating distributional issues (eg inequalities), and appropriately measuring the transfer of goods and assets across countries allowed by rapid technological advances.

Yet, the most acute challenges for statisticians stem perhaps from the rapid development of financial globalisation (BIS (2017)). For example, the global business model of multinational enterprises (MNEs)[7] makes it difficult to identify the location of corporate decision-making bodies – particularly in the case of **intellectual property products (IPPs)**, such as R&D and software products, which are recorded as intangible produced assets (2008 SNA, #10.98 and #A3.90). IPPs can be transferred across borders (in particular within a single multinational organisation) at low cost and with minimal business impact. The implications can be considerable, as seen in Ireland, where a number of multinational corporations attracted in large part by low corporation tax rates have relocated their economic activities, and more specifically their underlying intellectual property. This caused a large revision in Irish GDP in 2015, resulting in an impressive reported rate of growth (26%; OECD (2016)). But one difficulty related to such intra-group relocations of IPPs is to distinguish between legal and economic ownership, noting that the SNA recommends that assets be recorded on the balance sheets of the economic rather than the legal owner (2008 SNA, #2.47).

Another important issue is associated with the growing use of **special purpose entities (SPEs)**, which are legal entities typically used by companies as financial vehicles. This poses important challenges, especially in small open economies. One notable statistical caveat relates to the asymmetrical accounting treatment between resident firms and MNEs: in general, SPEs do not satisfy the definition of an

[7] See 2008 SNA, Chapter 21 (#21.47–48).

institutional unit in the SNA because they lack the ability to act independently from their parent corporation.[8] But when the parent company is a foreign entity, SPEs have to be recorded in line with the application of the domestic residency principles of the SNA – and this can be challenging for national accountants.

Several presentations at the conference touched on a particular aspect of these **broad statistical issues posed by the prevalence of multinational firms, the incentives to locate businesses in low-tax jurisdictions, and the growing use of offshore financing vehicles**. For instance, an estimation by the Federal Reserve Board showed that nearly one third of measured US cross-border portfolio investment can be distorted by standard reporting conventions (see Bertaut et al, session 1). Fortunately, it would seem that much work is going on to address such issues. For example, the Bank of France has developed a number of methodologies for estimating household portfolio investments in securities not covered by national reporting (see Gervais and Quang, session 1). Similarly, the Bank of Italy is "looking through" cross-border positions in investment funds so as to get a more representative description of the composition of portfolio investments (see Della Corte et al, session 3). Turning to the IMF, it has undertaken a long-term project to compile from-whom-to-whom[9] information for cross-border portfolio securities, taking advantage of its position as a central repository of international data (see Harutyunyan and Sánchez Muñoz, session 1). Furthermore, analysts at the Netherlands Bank are combining from-whom-to-whom data from financial accounts with information on international investment positions (IIPs) to build a sectoral network model for financial analysis (see Bijlsma et al, session 2).

One promising avenue seems to organise the **collection of globally integrated data sets** capturing group-level activities both within domestic jurisdictions and outside national borders. For instance, the Bank of Portugal has developed a business database which depicts the group structure of Portuguese firms and covers their resident and non-resident affiliates (see Pinto et al, session 2). Similarly, the Bank of France has designed a survey methodology to capture firms' involvement in the global economy, supporting the compilation of statistics on balance of payments (BOP) and IIPs (see Golfier, session 1).[10] Turning to the BIS, several initiatives have been launched to better understand the complex interconnections between entities located in different jurisdictions. One is to combine complementary statistical approaches to identify the residency of the ultimate holders of debt securities, which are sometimes difficult to allocate across countries (see He and Filkova, session 2). Another approach developed jointly by the BIS and the Bank of Portugal is to use so-called mirror data to improve information on the external sector in general and on cross-border banking claims (see Falcão Silva and Pradhan, session 1).

Filling data gaps does not require collecting new data in every instance. In some circumstances, unique combinations of existing data sets can be used to derive needed information. For example, one can mobilise publicly available information on

[8] See 2008 SNA Chapter 4, especially #4.55–56: "There is no common definition of an SPE but some of the following characteristics may apply (…). Such units often have no employees and no non-financial assets. They may have little physical presence" (…) and "are often resident in a territory other than the territory of residence of the related corporations".

[9] For an introduction to from-whom-to-whom tables in financial accounts, see Tissot (2016b).

[10] For a description of the BOP statistical framework, see IMF (2009), and for a review of current challenges related to external sector statistics, see IFC (2018a).

loans to construct a database on cross-country banking exposures (see Serena Garralda, session 2). Similarly, existing IIP data have been used at the Deutsche Bundesbank to calculate the impact of exchange rate changes on the domestic value of assets and liabilities for German residents (see Arz et al, session 1).

Non-bank financial intermediation

Another statistical shortcoming highlighted by the GFC was related to the activities of non-bank financial intermediaries, for which there was too little information available for policymakers despite their growing significance in the financial sector. These data needs were obvious from the perspectives of both monetary policy, especially to shed light on monetary transmission mechanisms, and financial stability, because of potential systemic risks building up in unregulated sectors that could spill over to the banking sector. Particular attention was paid to the shadow banking sector, which saw a substantial increase in its assets relative to those of banks in the run-up to the GFC.[11]

The conference provided ample evidence that the statistical community is developing **new data sets to better understand and measure the activities of non-bank financial institutions**. The Bank of Japan has recently compiled and released data on "Other Financial Corporations" – that is, the non-depository institutions which are often not subject to prudential requirements comparable with those applied to banks[12] – making full use of existing statistics including micro data sources (see Date et al, session 3). In Europe, the focus has been on greater data harmonisation and completeness, with ongoing work at the ECB to broaden the coverage and consider those financial institutions that are not already captured by existing standardised reporting schemes (see Agresti and Giron, session 3). More generally, many central banks have been investigating those particular aspects of non-bank financial intermediation that might contribute to specific financial stability concerns. One telling example was the research conducted at the Bank of Ireland related to the recent rise of Irish real estate investment funds and the implications of their activities (see McCarthy, session 3).

A number of presentations also shed light on **ongoing structural changes in the financial system**. One topic related to derivatives operations, in which non-financial corporations (and not just financial institutions) are increasingly involved. Micro, firm-level data can elucidate the characteristics of entities making use of derivatives, for instance, in terms of size, sector and financial structure. Going deeper and analysing even more granular data, one can assess firms' trading preferences in relation to different types of derivative contract using transaction-level information collected under the European Market Infrastructure Regulation (EMIR; see Benatti and Napolitano, session 3). A second important topic regarding the functioning of the financial system is related to the demand of currency in circulation. Some countries (eg Sweden) have seen a steady decline in the amount of cash transactions in recent

[11] See the FSB monitoring of non-bank financial entities' involvement in credit intermediation that could pose financial stability risks such as maturity/liquidity mismatches and excessive leverage (FSB (2018)).

[12] In the SNA framework, financial corporations are divided into three broad classes: financial intermediaries; financial auxiliaries; and other financial corporations (defined as "institutional units providing financial services, where most of their assets or liabilities are not available on open financial markets" (2008 SNA, #4.101)).

years, but there are many differences across countries, and paper-based payments such as cheques and cash still play important roles (Jakobsen (2018)). But assessing demand for cash is particularly challenging in a monetary union, and alternative approaches need to be followed (see Dias, session 4). Another issue discussed at the conference was Islamic finance, which can significantly alter financial intermediation patterns, for instance, by shifting the composition of banks' assets towards instruments financing physical assets and discouraging leverage creation. Yet the reporting practices of Islamic financial institutions vary substantially across jurisdictions, reflecting different types of business model, which calls for the development of more adequate international statistical guidelines – for instance, as regards the measurement of financial intermediation services and the characteristics of underlying Islamic financial products (see Goh, session 7).[13]

Collecting more granular data

Another lesson from the GFC was the need for more granular, institution-level data, and in particular for financial balance sheet information. Obviously, such data have been in demand to back up authorities in their resolve to tighten microprudential regulation and supervision after the crisis. However, they also help to shed light on the financial system more generally. In particular, there is a general push among central banks to **set up central credit registers (CCRs)**, ie a centralised system for collecting entity-level credit information on loans provided to the economy.[14] For instance, data from the Bank of Korea's consumer credit panel have proved to be particularly useful for supporting policy (see Kim, session 5). Similarly, the Bank of Israel has been developing a central credit registry with a proper disclosure control process (see Mantzura, session 7).

Such registries offer a wealth of information with many **practical applications**. For instance, the experience of the Central Bank of Malaysia is that granular credit data sources can be usefully combined to estimate the supply and demand components of household credit growth (see Soh, session 5).[15] Turning to the National Bank of the Republic of Macedonia, a new, granular credit registry data set is being used to investigate the existence and importance of a risk-taking transmission channel for monetary policy (see Miteski et al, session 4).

But the push for more granular data is **not limited to information on credit provided by banks**. At the ECB, for instance, in addition to the ongoing initiative to establish a European-wide CCR (AnaCredit),[16] important efforts have been made to set up detailed security-by-security data sets. As one of its benefits, this granular information permits an improved analysis of the portfolio rebalancing implications of the Eurosystem's asset purchase programme after the GFC, and the associated distributional effects across household groups (Kavonius and Honkkila, session 5). Another useful source of granular, firm-level information is provided by Central

[13] See 2008 SNA #6.163-6.169 for the issues related to the calculation of the financial intermediation services indirectly measured (FISIM).

[14] For a general discussion on central banks' use of CCRs, see IFC (2017c).

[15] The related paper from the Central Bank of Malaysia on "Disentangling the supply and demand factors of household credit in Malaysia: evidence from the credit register" (Soh (2018)) received the IFC award for the best paper presented at the conference by a young statistician.

[16] "AnaCredit" stands for analytical credit data sets.

Balance Sheet Data Offices (CBSOs), which generally provide a wealth of details covering firms' individual financial statements (IFC (2017b)). This can be useful for policy evaluation on how companies in different classes are affected by stress episodes (see Artman, session 5).

2. Turning data into information

As noted above, a second theme of the conference was the need to turn data into **information that is useful for policy**. To this end, statistical collection should go hand in hand with economic analysis: statisticians should strive to provide data that is suitable for the needs of analysts, who in turn should have the appropriate lens when interpreting the data collected. This entails two major consequences if one wants to fully reap the benefits of sound data. First, information has to be relevant for authorities, ie it can be fed into applications supporting policy analysis. From this perspective, various elements such as international harmonisation, high-quality standards, and user-friendly dissemination platforms can help to make the data more usable. Second, it should be possible to effectively use this information, that is, it has to be accessible to analysts and policymakers. This calls for proper arrangements for sharing data securely, taking into account confidentiality considerations. Needless to say, good communication between statisticians and other stakeholders is a fundamental building block to ensure progress on these two fronts.

Data for policy analysis

While it is of crucial importance that analysts apply an appropriate lens to the data, it is equally important that the data collected are tailored to their needs. Data requirements should thus be constantly evolving in response to the changing economic environment and new policy needs. This was clearly illustrated by Claudio Borio (BIS) in his keynote address, which discussed the **data requirements associated with evolving macroprudential policies**. A key factor has been the major intellectual shift that occurred after the GFC, with the realisation that episodes of financial instability can be the endogenous consequence of the previous build-up of macro-financial imbalances, and not just the result of exogenous shocks pushing the economy away from its long-term equilibrium.

This shift calls for new analytical frameworks, and in turn new types of data. For instance, granular, institution-level data are required for conducting macroeconomic stress tests and assess the potential resilience of financial institutions to adverse shocks as well as for evaluating the impact of new regulation. At the same time, aggregate measures remain indispensable for monitoring the financial system as a whole, for instance, to assess the stage of the credit cycle.[17] There is thus a clear need to "see the forest as well as the trees" (Borio (2013)), by combining micro and macro data sets (IFC (2016a)). One should note, however, that what constitutes **useful information to policymakers can vary over the financial cycle**: as emphasised by Agustín Carstens in his opening remarks, rough aggregates generally suffice to indicate that imbalances are building, but more granular data are needed for taking

[17] See, for instance, the various credit data sets developed by the BIS to this end, at www.bis.org/statistics/about_credit_stats.htm?m=6%7C380%7C673.

decisions once a crisis breaks out. One way of addressing these issues is to make sure that the new data frameworks are well integrated with central bank functions, such as microprudential regulation, macroprudential supervision of financial stability risks, crisis resolution etc.

The need to establish a link between the data collected and diverse policy needs was emphasised by many participants. For instance, analysis conducted at the BIS shows that CCR data can be useful for analysing both the impact of macroprudential policies and their interaction with monetary policy – two major issues for central banks in the aftermath of the GFC (see Gambacorta and Murcia, session 4). A number of other presentations highlighted the **vast range of policy needs that constantly call for developing new (or re-arranging existing) data sets**. In Brazil, where structural obstacles limiting the development of the domestic financial system are a primary policy concern, the central bank has developed a new set of indicators to support its efforts in lowering the cost of credit through various channels (see Fiorindo et al, session 4). At the ECB, a detailed security-by-security data set has shed additional light on the implications of the Eurosystem's asset purchase programme not only in terms of distributional effects, as noted above, but also as regards its impact on international capital flows (see Bergant and Schmitz, session 4). As another example, work at the central bank of Colombia demonstrates how unanticipated changes in the reference interest rate can affect the price of credit and saving instruments. To better understand the impact of monetary policy on various financial prices, "monetary policy shocks" are derived from a survey of experts' forecasts (and errors) of expected policy decisions (see Christiano-Botia et al, session 9).

Data-sharing and dissemination

Statisticians have faced a **difficult trade-off since the GFC**. On the one hand, there has been a steady demand for micro information. On the other hand, the more granular the data, the more challenging it is to address confidentiality concerns. Session 6 of the conference dealt with the issues posed by the management of granular financial data. This was a welcome opportunity to introduce the recently formed International Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA).[18] INEXDA aims to facilitate the international use of granular data for analytical, research and policy purposes, first by providing a forum for exchange of information and ideas on those issues and, second, by offering a framework for investigating possibilities to harmonise data access procedures for external stakeholders. An exchange of country experience has been awaited to address the statistical challenges associated with the GFC. To this end, INEXDA members have decided to develop a shared metadata scheme that can be applied to granular data sets from different countries (see Bender, Hausstein and Hirsch, session 6). This represents a useful first step in moving towards greater harmonisation and standardisation across countries as regards the procedures to access granular data from outside the central banks.

In parallel, important work has been going on to develop **technical solutions for sharing information** with external stakeholders while preserving anonymity and satisfying privacy protection laws. At the Bank of Italy, cryptographic techniques have been applied to link anonymised data sets (see Bruno et al, session 6). In particular,

[18] See Bender et al (2018).

four different types of data-sharing arrangement have been identified using these techniques. Addressing the same issue, but from a different perspective, various masking techniques have been developed at the Central Bank of the Republic of Turkey with the aim of providing user-friendly software tools to generate anonymised data sets. Work is under way to assess the impact of these techniques on the accuracy of the statistical operations conducted on such transformed data sets, as well as the degree of protection against potential attempts to uncover masked information (see Başer et al, session 6). Another example is the Bank of Israel's newly developed credit registry, which is supported by an in-house framework for anonymisation and control (see Mantzura, session 7).

Of course, the **challenges posed by sharing granular data** under confidentiality constraints do not relate only to external researchers, they are also relevant within central banks as well as between central banks and (domestic and foreign) authorities.[19] As regards the situation within central banks, a key issue is to establish a proper information management system to ensure that users have access to adequate data; this is indeed one important lesson of Bank of Mexico's 15 years of experience in collecting and sharing granular financial data (see Gaytan González et al, session 6). As regards sharing between central banks, a number of successful cooperative approaches have been implemented, such as the initiative by the Deutsche Bundesbank and the Bank of France for sharing data reported by complex multinational enterprises (see Mosquera Yon and Walter, session 5).

Moreover, confidentiality and privacy constraints are not the only factors that limit access to information. In order to be useful to analysts and policymakers, **data need to be clearly structured, presented in a manageable format, and communicated clearly** via a user-friendly system. To address these issues and improve its internal data-sharing capabilities effectively, the ECB has developed a Data Intelligence Service Centre (DISC). This central technology platform for organising, storing, and managing data is compatible with many common software tools for statistical analysis. The aim is to provide quick and easy access to data for internal users, including complex granular data sets (see Witt and Blaschke, session 7). The Deutsche Bundesbank's experience is that the success of such initiatives relies on setting up a well thought-out process for data integration to address the challenges posed by huge data volumes (see Müller, session 8) – noting that these data are neither well organised nor complete, with the increasing importance of "found" data, as distinct from "designed" data (that is, data collected for a specific statistical purpose).

From a similar perspective, but focusing more on the external users, the Reserve Bank of New Zealand and the Central Bank of the Republic of Turkey have explored user-friendly online graphical tools to **enhance the actual dissemination of publicly available central bank data**. In both cases, the general public can access tools to get "dashboard views", with customised reports and graphs, on an interactive basis. While for Turkey the system covers broad economic aggregates and is designed primarily to promote financial literacy (see Eken et al, session 7), the New Zealand tool focuses on the disclosure of banking statistics, disaggregated at the firm level, and is particularly designed to support market discipline in the banking sector (see Irrcher, session 7). One interesting issue relates to increasing demand for long consistent data

[19] On the general data-sharing issues faced by central banks, see IFC (2016b). For a recent example of international data-sharing related to the monitoring of global systemic institutions in the context of the DGI, see Bese Goksu and Tissot (2018).

series, while in reality break-adjusted series are rarely published, especially for data on stocks or positions. To address this need, the Netherlands Bank has introduced a policy to foster the compilation and publication of break-adjusted series, with adequate documentation on the methodology used (see van der Helm and Bartman, session 7).

3. New technological frontiers in statistics: big data, machine learning, and automation

Rapid technological change is **transforming central banks' internal statistical capabilities** (for instance, in storing and analysing data) **as well as data requirements** (for instance, to measure the development of fintech). As a result, big data and machine learning techniques have become increasingly relevant to the design of their information systems. Central bank statisticians are already using data derived from internet activities (eg Google searches) and applying machine learning techniques (eg principal component analysis, clustering). Yet interest in exploring these areas has clearly increased since the GFC, together with the actual capability to do so (thanks to increased computing power), and the availability of large data sets, both structured and unstructured. Indeed, in 2015 two thirds of the central banks surveyed by the IFC said they were actively discussing and investigating the use of big data, even though fewer than one third were already using big data sources in a meaningful way.[20]

As usual when innovation emerges, one risk is that these new techniques and data sources may disappoint those expecting major breakthroughs in official statistics (and for central bank information systems). Since more data are not necessarily better data, analysts and statisticians need to be aware of the limitations surrounding big data sets, especially as regards their accuracy and representativeness, which is sometimes overstated. **Understanding such limitations** is particularly important in the world of financial big data, where a clear feedback loop affects the information collected, the design of policy measures, and actions taken by market participants in response – implying that any move to measure a phenomenon can lead to a change in the underlying reality.[21] Furthermore, there are challenges associated with communicating results derived from "black-box calculations", especially for authorities who want to be transparent about their decision-making process. Hence, as emphasised by Professor Ruggeri (Vice President, International Statistical Institute (ISI)), one needs to be careful in using machine learning tools and to take **a balanced approach**. On the one hand, these techniques can be very useful in predicting and classifying, especially when one has to deal with a large number of parameters. On the other hand, they cannot solve all problems and it can sometimes be more appropriate to rely on a stochastic modelling approach to understand "what is going on". Addressing this trade-off requires a combination of these two types of technique; from this perspective, "statistical machine learning" would be a valid approach, but the range of solutions is still limited in practice.

[20] See IFC (2015b).

[21] Representing an application of the famous Lucas critique to the field of micro-financial statistics (Lucas (1976)).

Thus, a key message from the conference was that central bank analysts and statisticians **must cautiously explore the new opportunities** provided by technology innovation and be clear about the type of estimation being made, as well as its aims and limitations – in other words, "demystify" the concepts at stake (see Mehrhoff, session 8). At the same time, the conference was also an opportunity to present several **examples of big data applications in the central bank context**, suggesting that (i) there is a clear potential for big data sources to complement official statistics; and (ii) new, powerful statistical techniques can help users to navigate through the increasing amount of granular data collected since the GFC (IFC (2017a)).

As regards first the **potential for new data sources**, the possibility of applying indicators derived from internet activities in a central bank setting has already been well explored (Cœuré (2017)). For instance, the Deutsche Bundesbank is using Google data to produce a synthetic indicator for "nowcasting" mortgage market developments (see Oehler, session 8). "Financial big data sets" comprise not only the "internet of things", but also the wide range of granular data sets from commercial activities (eg credit card operations), financial markets (eg "tick-by-tick" price data), and administrative data sets (eg registers),[22] all of which can be usefully explored. One example from the Bank of Japan was related to detailed inter-dealer transactions data, which have helped new liquidity indicators to be constructed for the Japanese sovereign bond market (see Sakiyama and Kobayashi, session 8). Another area of interest relates to the development of text-mining techniques to access digitalised but unstructured information. For instance, the Bank of Canada has used a text analytics tool to develop a labour market conditions index for China, providing a useful complement to the (limited) official statistics (see Bailliu et al, session 8). Similarly, the Bank of Indonesia has developed an indicator based on text information produced by the media to assess general public expectations about policy rate decisions (see Andhika Zulen and Wibisono, session 9).

Turning to the second aspect, ie the **new tools provided by big data/machine learning/artificial intelligence technologies**, the central bank statistical community has shown a growing interest for them. One important application is to employ these new methods when the size of the data set is too big for "more conventional" tools. For instance, machine learning techniques have been used at the Bank of Greece to perform credit risk analysis derived from a very large loan-level data set (see Petropoulos et al, session 9).

Yet another promising avenue is to use these techniques to **improve existing data collections**. For instance, a BIS-developed algorithm for imputing missing data has proved capable of outperforming existing alternatives under certain conditions (see Kwon, session 9). Similarly, recent ECB work shows that techniques based on both supervised and unsupervised machine learning can be effective for outlier detection and missing data imputation (see Benatti, session 9). In addition, these new tools can support large-scale data cleaning exercises as currently undertaken at the ECB, for instance, to address the quality issues encountered with the collection of derivatives transactions reported by EU-based trade repositories (TRs). This is of particular interest because of the double-sided reporting requirements of the European EMIR Regulation: both counterparties of a trade are obliged to report this transaction to a

[22] For the use of administrative data sources for official statistics, see for instance Bean (2016) in the UK context.

TR, and statisticians manage data quality by comparing the resulting reports through an automated pairing and matching of the two-sided reports collected (see Pérez-Duarte and Skrzypczynski, session 9, as well as IFC (2018b)).

4. Conclusion: are post-crisis statistical initiatives complete?

The conference proved a **useful opportunity for taking stock** of the many statistical initiatives developed to address the information gaps identified in the wake of the GFC. Significant progress has already been made in terms of data collections, especially as regards the impact of globalisation on the financial system, activities related to non-bank financial intermediaries, and the need for more granular institution-level data especially on financial balance sheets. Yet the event was also a useful reminder that collecting data is a necessary but not a sufficient condition for knowledge. It is also crucial to turn data into information that is useful for policy, and this puts a premium on enhancing the dialogue between data compilers and analysts, especially within central banks. The goal is, first, to ensure that the statistics collected are effectively feeding into policy work, and, second, that the information is adequately shared with and disseminated to users. A third important message was related to the opportunities (and risks) associated with the rapid pace of technological change, especially since the GFC: new big data sources have emerged that can be of interest to central bank statisticians; and more sophisticated techniques are available to deal with the increasing amount of information at their disposal.

From this perspective, can one thus say that post-crisis statistical initiatives have been completed? The conference's concluding panel[23] showed that the **jury is still out**: a lot has been achieved over the past decade, but many more things remain to be done and this may perhaps take just as long. The following considerations were seen as particularly relevant for central bank statisticians looking ahead:

- **The requirement to fill known information blind spots cannot, by definition, be a finite exercise**. There is an ongoing, recurring cycle involving data compilation and data analysis, implying that statistical requirements will have to be continuously adapted to meet evolving policy needs. Cases in point relate to fintech and globalisation, two major sources of disruption in terms of future information requirements.
- Yet the **costs and benefits** of any new statistical initiative need to be carefully weighed up: data collections are, ultimately, burdensome exercises that warrant effective coordination with the stakeholders involved as well as careful long-term planning.
- Moreover, **statistics are not an end in themselves**. Collecting statistics is a necessary first step, but it is equally important to break down any barriers standing in the way of information-sharing and dissemination to users. This puts a premium on data standardisation, strong legal frameworks, anonymised data-sharing technologies and, most importantly, good communication both across country borders and between agencies working

[23] IFC Chair Claudia Buch was joined by Peter van de Ven (Head of National Accounts, OECD), Fabrizio Ruggeri (Vice President, International Statistical Institute), and Aurel Schubert (Vienna University of Economics and Business, and former Director General Statistics, ECB).

within the same jurisdictions. The ultimate goal should be to facilitate the integration of statistical work in policy analysis.

- **The future of central banks statistics depends on how the role of central banks will evolve**. From this perspective, policy data needs cannot be predicted with perfect foresight, which calls for central bank statisticians to be agile and responsive, while continuously seeking to produce statistics that are fit for purpose and aligned with users' needs.
- **International cooperation is crucial**. Improving the quality and use of available statistics is invaluable, through knowledge-sharing and the exchange of experiences. It is also necessary to ensure proper convergence across jurisdictions in terms of data collection and harmonisation, which is a key requirement given the global nature of the financial system.
- **There are clear potential benefits from ongoing technological innovation**. One should be aware of the risk of over-investing or adopting new tools for novelty's sake,[24] but central banks should actively explore the new opportunities offered by the big data revolution. Given limited resources, this also puts a premium on **cooperation both across countries as well as disciplines**. As a committee representing almost 100 jurisdictions that is associated with the ISI, the IFC can certainly help to promote such cross-country and cross-disciplinary fertilisation.

[24] Highlighting the importance for central banks of "looking beyond the hype", as can be argued in the case of cryptocurrencies (BIS (2018)).

References

Bank for International Settlements (BIS) (2017): 87th *Annual Report*, "Understanding globalisation", Chapter VI.

——— (2018): *Annual Economic Report*, "Cryptocurrencies: looking beyond the hype", Chapter V.

Bean, C (2016): *Independent review of UK economic statistics*, March.

Bender, S, C Hirsch, R Kirchner, O Bover, M Ortega, G D'Alessio, L Teles Dias, P Guimarães, R Lacroix, M Lyon and E Witt (2016): "INEXDA – the Granular Data Network", *IFC Working Papers*, No 18, October.

Bese Goksu, E and B Tissot (2018): "Monitoring systemic institutions for the analysis of micro-macro linkages and network effects", *Journal of Mathematics and Statistical Science*, vol 4, no 4, April.

Borio, C (2013): "The Great Financial Crisis: setting priorities for new statistics", *Journal of Banking Regulation*, vol 14, July, pp 306–17. Also published as *BIS Working Papers*, no 408, April.

Caruana, J (2017): "International financial crises: new understandings, new data", speech at the National Bank of Belgium, Brussels, February.

Cœuré, B (2017): "Policy analysis with big data", speech at the conference on "Economic and financial regulation in the era of big data", Bank of France, Paris, November.

European Commission, International Monetary Fund, Organisation for Economic Cooperation and Development, United Nations and World Bank (2009): *System of National Accounts 2008*.

Financial Stability Board (2017): *Financial Stability Implications from FinTech*, June.

——— (2018): *Global Shadow Banking Monitoring Report 2017*, March.

Financial Stability Board and International Monetary Fund (2009): *The financial crisis and information gaps*.

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

Irving Fisher Committee on Central Bank Statistics (IFC) (2015a): *Data-sharing: issues and good practices*, Report to BIS Governors prepared by the Task Force on Data Sharing, January.

——— (2015b): *Central banks' use of and interest in 'big data'*, IFC Report, October.

——— (2016a): "Combining micro and macro statistical data for financial stability analysis", *IFC Bulletin*, no 41, May.

——— (2016b): *The sharing of micro data – a central bank perspective*, IFC Report, December.

——— (2017a): "Big data", *IFC Bulletin*, no 44, September.

——— (2017b): "Uses of central balance sheet data offices' information", *IFC Bulletin*, no 45, October.

——— (2017c): "Data needs and statistics compilation for macroprudential analysis", *IFC Bulletin*, no 46, December.

——— (2018a): "External sector statistics: current issues and new challenges", *IFC Bulletin*, no 48, November.

——— (2018b): *Central banks and trade repositories derivatives data*, IFC Report, October.

International Monetary Fund (2009): *Balance of Payments and International Investment Position Manual* – Sixth Edition (BPM6).

Jakobsen, M (2018): "Payments are a-changin' but traditional means are still here", commentary on the Red Book statistics, available on webpage of the Committee on Payments and Market Infrastructures (CPMI).

Lucas, R (1976): "Econometric policy evaluation: A critique", *Carnegie-Rochester Conference Series on Public Policy*, vol 1, no 1, pp 19–46.

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

Shin, H S (2013): "The second phase of global liquidity and its impact on emerging economies", keynote address at Federal Reserve Bank of San Francisco Asia Economic Policy Conference, November.

Soh, J (2018): "Disentangling the supply and demand factors of household credit in Malaysia: evidence from the credit register", *IFC Working Papers*, no 17, October.

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

——— (2016b): "Development of financial sectoral accounts: new opportunities and challenges for supporting financial stability analysis", *IFC Working Papers*, no 15, November 2016.