Financial big data and policy work: opportunities and challenges

BRUNO TISSOT

2019 edition

STATISTICAL WORKING PAPERS

eurostat

Financial big data and policy work: opportunities and challenges

BRUNO TISSOT

2019 edition

Manuscript completed in January 2019.

Printed by the Publications Office in Luxembourg.

Luxembourg: Publications Office of the European Union, 2019

PDF: ISBN 978-92-79-99750-1 ISSN 2315-0807 doi:10.2785/093952 KS-TC-19-001-EN-N

Table of content

Introduction / Executive summary 4
1. Dual micro/macro dimension of financial statistics 5
1.1 Legacy of the Great Financial Crisis 5
1.2 Financial Big Data 5
2. Multiple opportunities: what Financial Big Data can bring 7
2.1 Macro-relevant micro information 7
2.2 Distribution information 7
2.3 Better macro aggregates 8
2.4 Policy design and assessment 8
2.5 New insights 9
3. Challenges 10
3.1 Variety 10
3.2 Complexity 10
3.3 Time dependency 11
4. Issues for financial authorities 12
4.1 Handling financial big data 12
4.2 Using financial big data 12
5. Selected experiences in dealing with large financial data sets 14
5.1 Derivatives transactions reported to Trade Repositories 14
5.2 Security-by-security data sets 15
6. References 17

Introduction / Executive summary

Public authorities working in the financial sphere have shown an increasing interest for big data. On the supply side, the amount of information at hand has boomed, reflecting not only the internet revolution, but also all the various initiatives launched after the financial crisis of 2007-09 to make a better use of existing sources of information – including, but not only, digital sources – and start new data collections, often on a very large scale.

As a result, “Financial Big Data” are quite specific. They primarily consist of the very large though relatively well-structured data sets derived from administrative and financial activities, for which new private data sources can play a major role.

Accessing these data can provide many opportunities for public authorities like central banks. But specific challenges also arise when handling these data and using them to support policy. In practice, this calls for being both pragmatic – ie starting with small pilot projects when exploring the financial big data universe – and ambitious – ie taking this opportunity to revisit existing information systems in a holistic way.

These various dilemma can be illustrated by two data collections of particular importance for central banks, one on derivatives transactions and one on the issuance of debt securities.

Author:
Bruno Tissot, Bank for International Settlements, Bruno.Tissot@bis.org

1 Dual micro/macro dimension of financial statistics

1.1 Legacy of the Great Financial Crisis

A key lesson from the Great Financial Crisis of 2007-09 (GFC) for central bank statisticians is that financial stability issues have both a micro and macro aspects. Entity-level developments can reverberate into the entire financial system, as was seen after the failure of one single financial institution (Lehmann Brothers) in September 2018. And, symmetrically, the situation of the financial system as a whole has certainly an impact on the individual situation of firms; for instance their financial performance depends on the evolution of global asset prices as well as on the stage of the financial cycle, especially whether it is in its upward phase (the booming years characterised by increased leveraging and risk-taking) or its downward phase (associated with widespread deleveraging and asset prices correction).

The consequence of this duality can be tricky for public agencies in general, and for central bank statistical units in particular. One needs to get a sense of developments at the level of individual firms or even of single transactions, and at the same time take due consideration of broader macrofinancial evolutions. In other words, “we need to see the forest as well as the trees within it” (Borio (2013)). The difficulty is that focussing on micro level data raises the risk of losing the “big picture”; conversely, relying only on macro aggregates can lead to miss important information that can be masked by average measures.

To be fair, a lot has been done, and is still being under way, to address this challenge. In particular, substantial progress has been achieved in the area of financial statistics with the implementation of the Data Gaps Initiative (DGI) endorsed by G20 authorities in the aftermath of the GFC. Indeed, the second phase of this DGI initiative, which covers the period between 2016 and 2021, explicitly calls for new statistical frameworks to combine micro- and macro-level data sets, a key objective being to “help straddle the divide between micro and macro analysis” (Financial Stability Board and International Monetary Fund (2015)).

1.2 Financial Big Data

The huge expansion in the micro information collected since the GFC has taken place in parallel to the “Big Data” revolution. Indeed, digital-based information offers great opportunities for official statisticians to supplant existing data sets or generate completely new statistical sources (MacFeely (2019)). Yet “Financial Big Data” is much more than just “Big Data” as commonly understood, especially when it relates to the daily work of central bank statisticians.

In practice, observers tend to refer to “Big Data” as the vast amount of information that has come to the fore due to the “internet of things” – ie the vast range of data created by connecting to the internet, with the expansion of the number and variety of devices that are able to collect information and communicate it (Lal Das et al (2017)). This includes data like the number of clicks on certain web

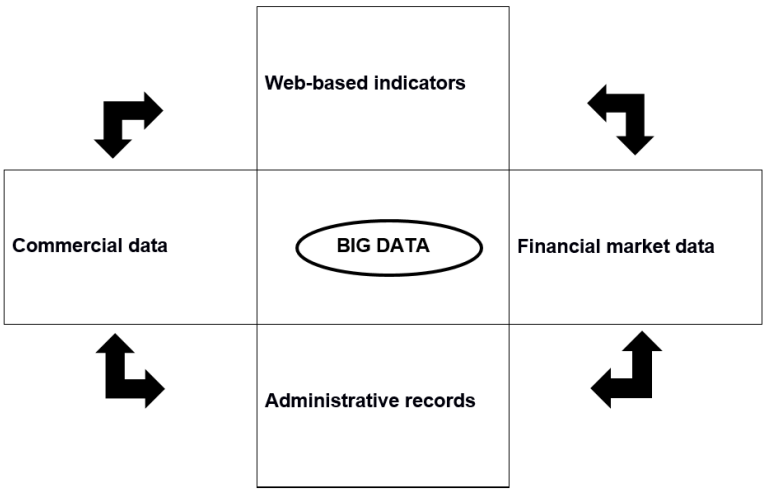
pages, the text contained in users' searches on websites, as well as all the signals sent by various types of sensors (mobile phones, satellites, pictures etc) through the web.

Indeed, public statisticians working on the financial system also use an increasing amount of information derived from the internet (Cœuré (2017)): Google searches, for instance to nowcast some macroeconomic indicators like retail sales; prices displayed on websites and that are extracted automatically – using "web scraping" techniques as with the Billion Prices Project developed at the Massachusetts Institute of Technology (MIT); (Cavallo and Rigobon, 2016)) – for instance to measure some components of inflation or residential price of houses; twitter messages, to assess the impact of changing sentiment among economic agents; etc. An important related development is the increasing use of textual information that is not strictly speaking derived from the internet but can be digitalised and, as such, analysed using big data-type text mining techniques.

Yet central banks' experiences show that "Financial Big Data" is far more than just the internet of things (Irving Fisher Committee on Central Bank Statistics (IFC) (2015)). The bulk of what their statisticians are dealing with on a daily basis consists of large, granular, and relatively well-structured data sets that are often a by-product of financial, commercial and administrative activities. Typically, these data are collected for some sort of "registers". This sort of data sets has clearly expanded since the GFC, reflecting both supply and demand factors. On the demand side, there has been a greater focus on micro data, as a way to go "beyond the aggregates", especially for policy purposes – as highlighted during the Eighth ECB Statistics Conference on "*Central Bank Statistics: moving beyond the aggregates*" in July 2016. On the supply side, public authorities in general, and central banks and financial supervisory bodies in particular, have started to collect more data to support their decisions. In addition, private commercial data providers have also clearly expanded their activities.

Reflecting these evolutions and the combined impact of internet/digitalisation, one can identify four main types of "Financial Big Data" sets (see Graph 1 and Tissot (2019)): administrative records (eg tax registers); financial data sets (eg "tick-by-tick prices" observed in financial markets); commercial records (eg credit card operations); and internet-type sources (eg Google searches).

Graph 1: Four main types of "Financial Big Data" sets



2 Multiple opportunities: what Financial Big Data can bring

The high granularity of financial big data sets presents a lot of advantages, as argued in greater length in Tissot (2018). To summarise, one advantage is the ability to “zoom in” and get information on very specific entity or transaction that is important from a system-wide perspective. A second benefit is to have a better sense of the distribution of macro aggregates. Third, the granular data points collected can, at least in theory, be assembled to get better “macro” estimates. Fourth, detailed information facilitates the design and evaluation of financial policies. Lastly, the new data sets can trigger new, innovative ways of looking at economic and financial phenomena.

2.1 Macro-relevant micro information

Collecting granular information on the financial system is clearly an advantage when one is interested in looking at the situation of individual entities that have a systemic feature. From this perspective, “Financial Big Data” provides a wealth, almost infinite source of potentially useful information. Indeed, the GFC emphasised the importance of understanding better what lies behind aggregates numbers; for instance, the situation of a systemic bank, which can be very different from the rest of the domestic banking sector. This information can be lost when looking only at “traditional” macro data sets. A key reason is the existence of non-linearity effects. For instance, exposures of individual entities trading with each other in a similar sector are “netted” when one aggregates these numbers at the sectoral level; yet in case of financial stress, the default of one entity can have spillover effects to its counterparts and lead to a systemic crisis.

A good example of increasing policy interest for collecting very granular data on systemic entities is provided by the International Data Hub hosted by the BIS (Bese Goksu and Tissot (2018)). The purpose is to share granular information so as to facilitate supervisors’ monitoring of global systemically important financial institutions (G-SIFIs). The focus is not on measuring the average situation of G-SIFIs together, but rather to extract micro information important for macro financial stability work. A similar approach is also being developed for other segments of the financial system, for instance to detect potential stress points related to repos or derivatives transactions – with the need to deal with very large data sets (see below).

2.2 Distribution information

A second benefit of granular data sets is to measure and take into consideration the distribution of indicators in the population of interest. Indeed, “Financial Big Data” allows for a better exploring of the heterogeneity behind aggregated numbers and for conducting so-called “tail analysis”. For instance, the GFC showed how important it was to have a good understanding of indebted homeowners with relatively poor credit quality (the US subprime mortgage market); this specific

information, and related underlying fragilities, was clearly different from the aggregate indebtedness situation measured at the overall level of the households sector.

The response to the GFC also underlined the importance of distribution aspects for policy actions. For instance, there has been a significant push to develop so-called macroprudential tools to better address financial stability risks (cf Borio (2018) and IFC (2017a)). By nature, applying such tools requires having a sub-macro approach. For instance, measures to restrict credit (eg loan-to-value ratios) can be implemented in different ways depending on the sector considered (eg consumer credit versus mortgage credit), specific debtor groups (eg first-home buyers versus commercial property developers) and even instruments (eg fixed interest rates versus floating interest rates). In other words, the tools cannot be identical for the economy as a whole, when some differentiation is required to target specific segments of the financial system.

2.3 Better macro aggregates

Financial big data sets can also be very useful for the compilation of macro statistics, a bit like when one assembles "Lego bricks". In particular, micro, well-structured "administrative" data sets can be quite helpful in this endeavour, especially in comparison to other statistical sources (Bean (2016)). For instance, a number of macro indicators are compiled based on surveys, whose coverage of the entire population of interest can be limited; using more comprehensive and granular data sets can then enhance the calculation of aggregated indicators. Another benefit reflects the fact that many of the granular data sets available are "found data", ie they exist independently of any statistical production, for instance because they are the by-product of an administrative or commercial activity. Since the data already exist, the additional collecting cost for statisticians is limited (and so is the associated reporting burden for economic agents). In contrast, setting up a dedicated statistical survey/census can be costly and burdensome.

As regards financial statistics in particular, several micro data sets appear useful in order to enhance the computation of "macro" aggregates (IFC (2016a)). Many central banks have stepped up their efforts since the GFC to develop large and consistent loan-by-loan and security-by-security data sets, both in terms of transactions (eg issuance of new loans and securities) and stocks (eg holdings of financial instruments). Another source of granular, firm-level information is provided by Central Balance Sheet Data Offices (CBSOs), which generally provide a wealth of details covering firms' individual financial statements (IFC (2017b)). All these data sets have proved particularly useful for those in charge of compiling the financial accounts of the economy, because they provide an extensive coverage of the financial transactions and positions of the various types of economic agents, and also because they facilitate the estimations of the financial linkages between various sectors (eg computation of so-called "*from-whom-to-whom*" tables; cf van de Ven and Fano (2017)). At the same time, these data sets, being very granular, can be flexibly combined depending on actual analytical needs.

2.4 Policy design and assessment

The post-GFC period has seen an increasing interest for using financial big data sets for policy purposes, especially when designing, calibrating, assessing and modifying policy actions. This is obviously important for micro-prudential authorities, which have a keen interest in monitoring the implementation of the financial reforms by supervised entities, and the impact of these reforms on their business. In particular, so-called "Quantitative Impact Studies" (QIS) have now become a central element of the new indicator-based frameworks developed to, among other tasks, draw the lessons of previous policies, assess the ex-ante impact of new measures, identify additional areas of weakness, and clarify the functioning of policy – with the leading contribution made by the Basel Committee on Banking Supervision (BCBS) in this area (BCBS (2018)).

Yet a much wider range of public authorities are increasingly interested in the information content of

micro data sets for designing or applying public policies, even at a macro level – for instance to conduct macro prudential, monetary, fiscal and even structural policies (Tissot (2017)). A key factor is that the granularity of the data sets helps to capture very distinct, specific dimensions that are relevant for the policy actions considered (eg when the purpose is to modify the behaviour of targeted economic agents). In addition, the data can be used by multiple users: for instance, payments data can be interesting for those supervising the country's payments infrastructure as well as by those compiling BoP statistics (eg to measure tourism activities). Lastly, the granularity of the information collected, both at a point in time and across time, helps to understand feedback effects, behavioural responses, unintended consequences, as well as cross-sectoral impacts…

2.5 New insights

Lastly, a key benefit of financial big data sets is their analytical potential. First, they provide a very rich view of the population of interest: for instance, all the securities issued in a country, by sector, instrument, currency, maturity, etc. Second, the data are often collected over a long period of time, especially when they serve a commercial or administrative purpose; this can be very helpful for analysing the evolution of the financial system over a long-time horizon, in turn providing useful insights looking ahead. Third, while public authorities are often just at the beginning of making sense of such large data sets, they can benefit from the practice accumulated by the private sector in this area. For instance, the experience of commercial banks in conducting stress test exercises using the information provided by their granular business records has become a notable source of inspiration for authorities monitoring potential fragilities in the financial system.

But perhaps more importantly, the greater availability of granular information offers new possibilities for analysis. Obviously, entity-level granular datasets can be used for analysing heterogeneity across economic agents as well as to conduct more detailed and varied types of data aggregation (eg by combining multiple criteria crossed, like firm size, sector, location etc; cf Pinto Ribeiro et al (2010)) – in itself pushing further the knowledge frontier.

One can even argue that this can act as a trigger for re-thinking and for adopting a different mind-set when analysing the economic and financial system. A particularly relevant example is the ability to use very detailed firm-level information collected on a residency-basis – in line with the System of National Accounts framework (SNA; European Commission et al (2009)) – but which can be recombined to provide aggregates that are consistent with the structure of corporate groups. For instance, the banking statistics of the BIS (McGuire and Wooldridge (2005)) are presented both from a "locational" perspective (that is, by summing up for all countries the banks that are resident in the country, even if they are affiliates of foreign groups) and a "consolidated" perspective (that is, by aggregating all the banking affiliates, resident or not, that are controlled by the "national" banking groups established in the country). Similar work has been also conducted using debt securities statistics (see below). Combining residency-based and nationality-based approaches has proved particularly useful to understand the functioning of the global financial system and the development of financial stress, for instance during the GFC (IAG (2015)).

3 Challenges

Despite many benefits, financial big data sets also present important challenges for central bank statisticians. Three important points are worth highlighting from this perspective: the challenges posed by the variety of the data sets, by their complexity, and by time-dependency issues.

3.1 Variety

In practice, financial big data sets comprise a large variety of heterogeneous information. Statisticians will therefore have to deal with a multitude of data formats, sources etc when trying to make use of the information at hand. This difficulty is reinforced by the fact that most of the data considered are “found” data, ie they are the by-product of non-statistical activities. This contrasts with more “traditional” statistical data sets (eg surveys, census etc), which are “designed data”, explicitly collected for pre-defined statistical purposes.

This variety raises several issues for public statisticians, including the costs of either getting the data (esp. when they are sold by commercial vendors) or compiling them (eg costs associated with cleaning the various data sets and for compiling something that makes sense). Another problem is the quality of the data: this quality may not be sufficient because this is not a key element of the collection process itself, or because the data sets are simply too large (existence of missing values), or because – even if the data sets are large – they are not representative of the underlying population.

Yet a last, often cited problem relates to the difficulty to “link” the various data sets. For instance, one may want to merge the information provided on loans with information on debt securities issuance. But for doing that one needs to have correct identifiers to connect the data points. The situation is even more difficult when looking at corporate groups, which can comprise a vast range of affiliates; in the absence of explicit parent relationships between these entities, the aggregation of this information may prove uneasy because of the difficulty to identify groups’ structures – an issue that the international community is trying to address with the Legal Entity Identifier (LEI) initiative (LEIROC (2016)).

3.2 Complexity

Complexity is a second key challenge. The micro-level data universe is complex and constantly evolving depending on the evolution of the financial system, which is changing at a very fast pace compared to the “more traditional” real economic sphere. Second, the use of a particular data set may depend on the issue of interest: for instance, statisticians will be interested only in specific parts of the information provided by payments transactions, depending on whether the purpose is to monitor payments infrastructures or to measure the impact of tourism (eg cross-border transactions).

Addressing this complexity requires transforming raw, “organic” data into “smart data”, ie information that is relevant for policy. From this perspective, public authorities may just be at the beginning of this process. While most post-GFC efforts were devoted to “*collecting data points*”, the focus is now more on “*connecting the dots*” (Caruana (2017)).

3.3 Time dependency

A third main challenge relates to time dependency. Indeed, information needs evolve over times, so statisticians have to adapt data collections to evolving analytical purposes. This is particularly obvious for the information related to the financial system, characterised by its rapid pace of innovation and the importance of feedback effects: many financial data are collected to guide policy actions, which are precisely designed to influence the behaviours of economic agents that the statistical apparatus tries to capture. For instance, supervisors collect information on the exposures of regulated banks, on whose base they can decide to require a change in business models (eg greater capital accumulation and less leverage). This can trigger significant changes in the services provided by the financial industry (eg reduced lending to the economy), and in turn in the type of statistics that need to be collected (eg increased importance of non-bank financial intermediation compared to traditional lending; FSB (2018)).

An important aspect of these time dependency challenges relates to the fact that statistical needs depend on the state of the financial cycle. Before a financial crisis, it may be very difficult to spot precisely where pressure points are developing; assessing how fragilities are building up in the financial system will typically rely on aggregated statistics to spot “abnormal patterns” – say, rapid rate of credit creation, surge in cross-border lending, etc. In contrast, once the crisis hits, resolution work will usually request much more timely and precise information; this will put a premium on collecting information at the level of specific institutions or transactions (Carstens (2018)). Bottom line is that the demand for data is time-dependant: statistics on “rough aggregates” would be sufficient in “good times”, but one would need much more granular data in “bad times”.

4 Issues for financial authorities

The challenges posed by financial big data – their variety, their complexity, and time dependency aspects – raise very specific issues for public authorities like central banks, especially when handling and using these data. In practice, this calls for being both pragmatic – ie starting with small pilot projects when exploring the big data universe – and ambitious – ie taking this opportunity to revisit existing information systems in a holistic way.

4.1 Handling financial big data

Handling financial big data poses of course the question of resources. As always, money is the nerve of war – to acquire up-to-date IT equipment, security tools, specific staff skills etc. But this may not be sufficient, and these resources can easily be wasted if they are not used efficiently. What is needed is to set up proper arrangements for managing this information, including adequate statistical production chains to extract relevant statistics from (often noisy) data points. This calls for having a comprehensive information management process, with proper governance rules – "*who is in charge of what, and how is this done?*".

One could argue that these issues are relevant not only for financial authorities but for all entities working with large statistical data sets. Yet something that may be very specific to independent institutions like central banks is reputation. By nature, central banks require public trust to perform their tasks – issuing currency, conducting monetary policy, promoting financial stability, monitoring financial institutions (for those central banks tasked with micro prudential supervision). They are therefore expected to set up adequate statistical processes when dealing with financial big data, in a sense "leading by example" – ie being transparent about the purpose of the data collections; ensuring that the statistics are collected in line with strong ethical standards; taking adequate measures to preserve the information; etc.

This means, in particular, that economic agents have to be made fully aware of what data are collected and for which purposes… something that is not common practice in the commercial big data universe, despite increased public interest to address such issues – cf the introduction of the European General Data Protection Regulation (GDPR) in 2018. Reporting institutions and households alike need also to be reassured about the protection of the information collected, in particular as regards confidentiality rules and cyber security risks, even though this has to be properly balanced with the needs for greater data sharing across authorities (IFC (2016b)).

4.2 Using financial big data

Turning to using financial big data, a key problem is their degree of accuracy. In practice, one will tend to consider that, because the related data sets are very large, they provide a reliable description of the economic and financial reality. But this is not certain at all, and much depends on the coverage

bias of these data sets, which is often unknown. For instance, is it reliable to rely on statistics gathered on the web for those countries where a sizeable part of the population has no access to the internet? What is the information relevance of text messages collected on social medias (eg Twitter) if important groups of the population are not using these medias? Etc. In fact, even very large big data samples could be much less accurate compared to (smaller) traditional probabilistic samples (Meng (2014)). Central banks need therefore to be transparent about the limitations (eg representativeness, accuracy, quality) of the data used in the context of their decision-making process. Important from this perspective are the associated reputation aspects as well as the potentially high social costs of misguided policy decisions that would be based on "wrong" information.

A second type of issues when using financial big data is how they can influence the conduct of policy. One risk is that the availability of more, high-frequency data can develop a bias towards responding to news and encouraging shorter horizons. Furthermore, the greater ability to assess economic agents' sentiment through the use of social medias or text mining techniques may well lead to a greater focus on fine-tuning policy communication, with the risk of paying more attention to the impact of policy decisions on agents' perceptions compared to the underlying reality. Lastly, communicating policy decisions based on large data sets and "black box" calculations may disturb economic agents interested in getting "factual evidence". Whether central banks will be able in a not too distant future to explain their policy actions by using "soft signals" derived from opaque machine learning techniques and social medias (instead, as today, of "real" indicators such as output growth and inflation) thus remains to be seen.

5 Selected experiences in dealing with large financial data sets

Public authorities in general, and central banks in particular, are certainly just at the beginning of making use of "financial big data", despite the ongoing increase in projects being launched (IFC (2017c)). The points above suggest that this journey may present significant difficulties but also provide many opportunities, a duality illustrated by two recent developments: the initiative to collect transaction records in derivatives markets; and the growing interest among central banks for collecting and making use of security-by-security data.

5.1 Derivatives transactions reported to Trade Repositories

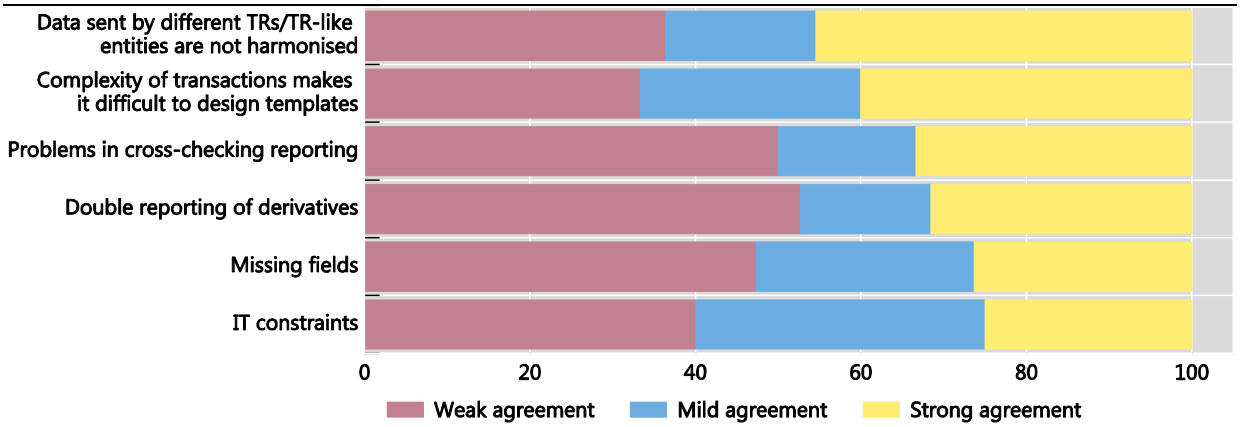
One important consequence for financial statistics of the reforms undertaken after the GFC has been the collection of a very large amount of data on activity in the derivatives markets (FSB (2010)). This resulted from the decision to request transactions in over-the-counter (OTC) derivatives markets to be reported to trade repositories (TRs). A key issue for public authorities now is to ensure that these data are effectively used, not least to guide policy actions.

From this perspective, a recent survey of central banks shows that the statistical coverage of the main OTC derivatives market segments has greatly improved (IFC (2018)). Yet significant data gaps remain, in particular to track transactions between unregulated entities; there is also room for improvement as regards critical derivatives data elements such as notional value, clearing information and termination dates. But a major challenge is the limited quality of the data collected, in particular when dealing with transactions that are reported to different TRs and with different attribute values. These difficulties are reinforced by the extremely large size of the related data sets, making human intervention to "clean the data" almost impossible, and calling for automatised approaches (eg machine learning; Cagala (2017)).

Moreover, better data quality is a necessary but not sufficient condition for supporting policy use. Authorities need to be able to make sense of the vast amount of the micro-level information collected, in particular by compiling summary indicators (see Fache Rousová et al (2016) for a review of the European experience for derivatives data). According to the central banks surveyed by the IFC in 2018, data aggregation has become a particularly important issue when processing this information, not least because transaction records are not harmonised among TRs (Graph 2). In addition, the sheer complexity of derivatives transactions makes it difficult to design appropriate statistical reporting templates. Furthermore, the lack of adequate identifiers as well as legal constrains are significant obstacles to deal with, especially because the trades can be reported by different TRs located in various jurisdictions. Other problems relate to the double reporting of transactions, which is difficult to correct, and missing attributes.

Graph 2: Issues related to the aggregation of granular derivatives statistics

Does your central bank face problems in the aggregation process of data collected by TRs/TR-like entities?



Source: IFC survey on TR data, 2018 (IFC (2018)).

5.2 Security-by-security data sets

A growing number of central banks have embarked on collecting security-by-security information especially since the GFC. The BIS has a specific experience in this area, as it publishes global statistics on debt securities issuance in international markets (that is, the markets where non-residents issue). To this end, the BIS draws on very granular information provided by commercial data vendors.

An important task is to get through this vast amount of information and classify the respective securities. In particular, the BIS follows specific criteria to determine whether a security can be considered as an international security or not. For each security, the issuer's country of residence is compared to three other locations (of the registration domain, of the listing place and of the law governing the bond). The security is thus classified as "international" if one of these three locations differs with the country of residence of the issuer. In reality, however, the characteristics of the data sets may not provide consistent information, and for a significant part of the exercise the result may not be conclusive. In addition, the importance of these quality issues will depend on the specific commercial data source used and of the period of time considered.

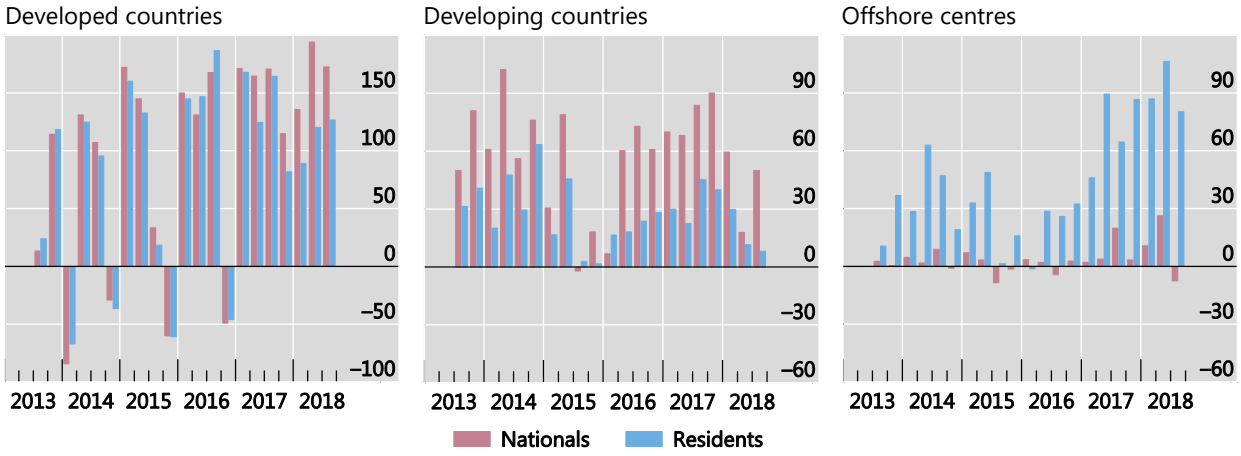
Despite these challenges, the BIS international debt securities data set provides many analytical opportunities. Obviously, because the information is highly granular, one can compute various aggregations – gross/net issues, repayments, amounts outstanding – with multiple breakdowns: by instrument, currency, maturity bands, interest rate, rating, guarantees etc. This provides a lot of flexibility in addressing a wide range of policy questions.

In particular, it is possible to classify and aggregate international debt issuance by the residency and the nationality (defined as the residency of the controlling parent) of the issuer (Gruić and Wooldridge (2015)). While the residency-based aggregates provide a picture that is consistent with the SNA / BoP framework, the nationality-based ones help to understand the consolidated exposures taken by issuers borrowing through their foreign affiliates (Tissot (2016)). This has been particularly useful in the recent years marked by a very large expansion of the debt issued in offshore centres by corporate groups from emerging market economies (EMEs), as part of the global shift to financing via bond markets and non-bank financial intermediaries (Shin (2013)). In fact, residency-based statistics of debt issuance in offshore centres has been inflated by the increased activity of foreign affiliates located in these centres but controlled by parent groups located in EMEs. Moving to

nationality-based aggregates helps therefore to a have a more comprehensive view of the consolidated exposures of EMEs firms issuing either directly or through their foreign affiliates (Graph 3).

Graph 3: International debt securities issued by financial and non-financial corporations[1]

Net issuance by region, in billions of US dollars[2]



Further information is available at www.bis.org/statistics/secstats.htm.

[1] Excluding general government. [2] For a list of countries in each region, see http://stats.bis.org/statx/srs/table/c1.

Sources: Dealogic; Euroclear; Thomson Reuters; Xtrakter Ltd; BIS debt securities statistics.

6 References

Basel Committee on Banking Supervision (BCBS) (2018): *Basel III Monitoring Report*, October.

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

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.

——— (2018): "Macroprudential frameworks: experience, prospects and a way forward", speech on the occasion of the Bank's Annual General Meeting, Basel, June.

Cagala, T (2017): "Improving data quality and closing data gaps with machine learning", in *IFC Bulletin*, no 46, December.

Carstens, A (2018): "Data for financial stability: collecting and connecting the dots", Remarks at the Ninth European Central Bank Statistics Conference, July.

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

Cavallo, A, and Rigobon, R (2016): *The Billion Prices Project: Using Online Prices for Measurement and Research*, Journal of Economic Perspectives, Spring 2016, Vol 30(2): 151-78.

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*.

Fache Rousová, L, K Kulmala and M Osiewicz (2016): "Reporting of derivative transactions in Europe", in *IFC Bulletin*, no 41, May.

Financial Stability Board (FSB) (2010): *Implementing OTC Derivative Market Reforms*, October.

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

Financial Stability Board and International Monetary Fund (2015): *The financial crisis and information gaps – Sixth Implementation Progress Report of the G20 Data Gaps Initiative*.

Gruić, B and P Wooldridge (2015): "BIS debt securities statistics: a comparison of nationality data with external debt statistics", in *IFC Bulletin*, no 39, April.

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.

Irving Fisher Committee on Central Bank Statistics (IFC) (2015): *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): "Data needs and statistics compilation for macroprudential analysis", *IFC Bulletin*, no 46, December.

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

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

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

Lal Das, P, S C Beisswenger, S Mangalam, M R Mehmet Rasit and M Lukac (2017): *Internet of things: the new government to business platform - a review of opportunities, practices, and challenges*. Washington, D.C.: World Bank Group.

Legal Entity Identifier Regulatory Oversight Committee (LEIROC) (2016): *Collecting data on direct and ultimate parents of legal entities in the Global LEI System* – Phase 1, 10 March.

MacFeely, S (2019): "Big Data and Official Statistics". In S Strydom, & M Strydom (eds.), *Big Data Governance and Perspectives in Knowledge Management* (pp. 25-54). Hershey, PA: IGI Global.

McGuire, P and P Wooldridge (2005): "The BIS consolidated banking statistics: structure, uses and recent enhancements", *BIS Quarterly Review*, September, pp 73–86.

Meng, X (2014): "A trio of inference problems that could win you a Nobel Prize in statistics (if you help fund it)", in X Lin, C Genest, D Banks, G Molenberghs, D Scott and J-L Wang (eds), *Past, present, and future of statistical science*, Chapman and Hall, 2014, pp 537–62.

Pinto Ribeiro, S, S Menghinello and K De Backer (2010), "The OECD ORBIS Database: Responding to the Need for Firm-Level Micro-Data in the OECD", *OECD Statistics Working Papers*, 2010/01, OECD Publishing, Paris.

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.

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): *Using micro data to support evidence-based policy*, International Statistical Institute 61st World Statistics Congress, July.

——— (2018): "Providing comparable information to assess global financial stability risks", *Eurostat Statistical Reports*, January.

——— (2019): "Making the Most of Big Data for Financial Stability Purposes", in S Strydom, & M Strydom (eds.), *Big Data Governance and Perspectives in Knowledge Management* (pp. 1-24). Hershey, PA: IGI Global.

van de Ven, P and D Fano (eds.) (2017): *Understanding Financial Accounts*, OECD Publishing, Paris.

Getting in touch with the EU

In person
All over the European Union there are hundreds of Europe Direct Information Centres. You can find the address of the centre nearest you at: https://europa.eu/contact

On the phone or by e-mail
Europe Direct is a service that answers your questions about the European Union. You can contact this service
– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
– at the following standard number: +32 22999696 or
– by electronic mail via: https://europa.eu/contact

Finding information about the EU

Online
Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu

EU Publications
You can download or order free and priced EU publications from EU Bookshop at: http://bookshop.europa.eu. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/contact)

EU law and related documents
For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex at: https://eur-lex.europa.eu

Open data from the EU
The EU Open Data Portal (http://data.europa.eu/euodp/en/data) provides access to datasets from the EU. Data can be downloaded and reused for free, both for commercial and non-commercial purposes.

KS-TC-19-001-EN-N

Financial big data and policy work: opportunities and challenges

Public authorities working in the financial sphere have shown an increasing interest for big data. On the supply side, the amount of information at hand has boomed, reflecting not only the internet revolution, but also all the various initiatives launched after the financial crisis of 2007-09 to make a better use of existing sources of information – including, but not only, digital sources – and start new data collections, often on a very large scale.

As a result, "Financial Big Data" are quite specific. They primarily consist of the very large though relatively well-structured data sets derived from administrative and financial activities, for which new private data sources can play a major role.

Accessing these data can provide many opportunities for public authorities like central banks. But specific challenges also arise when handling these data and using them to support policy. In practice, this calls for being both pragmatic – ie starting with small pilot projects when exploring the financial big data universe – and ambitious – ie taking this opportunity to revisit existing information systems in a holistic way.

These various dilemma can be illustrated by two data collections of particular importance for central banks, one on derivatives transactions and one on the issuance of debt securities.

For more information
https://ec.europa.eu/eurostat/

Publications Office

ISBN 978-92-79-99750-1