Irving Fisher Committee on Central Bank Statistics

IFC Bulletin

No 53

Micro data for the macro world

Proceedings of the IFC-CBM-ECB Satellite Seminar, Kuala Lumpur, Malaysia, 16 August 2019

April 2021

BANK FOR INTERNATIONAL SETTLEMENTS

Contributions in this volume were prepared for the proceedings of the IFC Satellite Seminar on "*Micro data for the macro world*", co-organised with the Central Bank of Malaysia and the European Central Bank, in Kuala Lumpur, Malaysia, on 16 August 2019.

The views expressed are those of the authors and do not necessarily reflect the views of the IFC, its members, the BIS and the institutions represented at the meeting.

This publication is available on the BIS website (www.bis.org).

ISSN 1991-7511 (online)

ISBN 978-92-9259-469-5 (online)

Micro data for the macro world

IFC Bulletin No 53
April 2021

Proceedings of the IFC Satellite Seminar on "*Post-crisis data landscape: micro data for the macro world*", co-organised with the Central Bank of Malaysia (CBM) and the European Central Bank (ECB)

Kuala Lumpur, Malaysia, 16 August 2019

Seminar overview

1. New frameworks for collecting micro data

2. Accessing and using micro data sets

Incorporating micro data into macro policy decision-making

Jean-Marc Israël and Bruno Tissot[1]

Executive summary[2]

In 2019, the Irving Fisher Committee on Central Bank Statistics (IFC) of the Bank for International Settlements (BIS), together with the European Central Bank (ECB) and the Central Bank of Malaysia, decided to analyse the **issues posed by the integration of micro data sets into the framework of official statistics, for the purposes of policymaking**.

This was an opportunity to review the expansion of micro (granular) data sets made available since the Great Financial Crisis (GFC) of 2007–09 and their possible contribution to macro compilation exercises. The review underlined the **importance of broadening official statistical frameworks to benefit from the growing availability of granular data** to assist the assembling of macroeconomic aggregates and/or facilitate the linkage between micro- and macro-level statistics.

This stocktaking exercise also documented that the **"micro data revolution" can bring important analytical benefits and effectively support central bank policies**, with a greater ability to "zoom in" on particular areas of interest and to assess the distribution of macroeconomic aggregates within reporting populations. One significant example is that granular and micro data already assist in assessing how finance can contribute to the greening of economies.[3] The initiative also proved a welcome opportunity to highlight the potential of new types of granular information in "unusual" circumstances. Indeed, and as observed subsequently with the sudden outbreak of the Covid-19 pandemic in early 2020, micro data sources can be a valuable complement to the "traditional" official macro statistics available in times of stress, especially to alleviate compilation disruptions, assess pressure points, and facilitate the implementation of targeted policy measures.

However, there are **also important challenges associated with dealing with these data sets, for instance as regards their quality, confidentiality and manner of accessing them**. Indeed, the task of integrating (granular and) micro financial information in macro frameworks has proved more complex than initially thought. To ensure concrete progress in the future, attention should focus on:

1 Respectively, former Head of the Monetary and Financial Statistics Division at the ECB (jeanmarc.il@free.fr); and Head of Statistics & Research Support, BIS, and Head of the Secretariat of the Irving Fisher Committee on Central Bank Statistics (IFC) (Bruno.Tissot@bis.org).

2 The views expressed here are those of the authors and do not necessarily reflect those of the BIS, the Central Bank of Malaysia, the ECB, the IFC or any of those institutions represented at the seminar. We thank Barend de Beer, Neil Humphries, Robert Kirchner, Michael Kock, Lim Sheng Ling, Ong Li Ming, Eric Nielsen, Laurent Olislager, Steffi Schuster and Johannes Turner for helpful comments and suggestions.

3 See eg the OECD's Microdata Analysis for Environmental Finance and Investment (MEFI) initiative.

(i) building effective micro data collection frameworks based on a comprehensive data strategy helping to contain reporting burden;

(ii) accessing and making use of more granular sources of information, with the need to overcome the challenges related to their size and complexity with a view to transforming simple data points into knowledge;

(iii) promoting the exchange of experience, eg as regards access to micro data sets and external research projects, the development of diversified staff skills, and the combination of different types of data sets;

(iv) developing new and adequate analytical tools, for instance to enhance data quality assurance processes, extract summary indicators from a wealth of data points, and develop machine learning (ML) / text mining / network analysis approaches to maximise the potential of granular (including micro-level) information; and

(v) bridging the gap between micro- and macro-level statistical exercises, which can be instrumental in enhancing the understanding of how the financial system functions and interacts with the economy, assessing distributional issues and facilitating sectoral analyses.

In supporting the above tasks, **there is a growing interest in making use of granular information from private sources that are not part of the official statistical offering**. What is unclear, though, is how data producers located outside the national statistical system should feature vis-à-vis the fundamental principles that govern the production of appropriate and reliable official statistics and adhere to certain professional and scientific standards. Helpfully, a number of private firms have already adopted dedicated and transparent mission statements and principles to address these concerns, and such initiatives may deserve to be strongly encouraged by the official (international) statistical community.

Lastly, there are **specific communication challenges, as well as potential legal difficulties, for policymaking institutions** like central banks when (confidential) granular analytical insights are used as the foundation for their decisions.

Introduction

Increasingly vast and complex amounts of granular data are becoming available for public authorities to use. This is particularly the case for central banks in their role as statistics compilers, a function that has clearly expanded in recent decades, reflecting the growing importance of finance in the economy (see Greenwood and Scharfstein (2013) for the US situation). Another trigger was the launch of various and important micro data collections, especially in the context of the G20-endorsed Data Gaps Initiative (DGI) in response to GFC. In particular, the first phase of the DGI (2009–15) stressed the importance of complementing macro-level aggregates with distributional information (see FSB and IMF (2009), Recommendation no 15). The second phase (2016–21) called for collecting more granular data to "help straddle the divide between micro and macro analysis" (FSB and IMF (2015)). Such an emphasis on the importance of micro data was relatively new in the area of official statistics. For instance, the last version of the Manual for the System of National Accounts (SNA) devotes only a two-page section (out of almost 700 pages) to "The use of microdata for macroeconomic accounting" – stating that "there would be considerable analytical advantages in having micro databases that are fully compatible with the corresponding macroeconomic accounts" but that this task "may be difficult, if not impossible, to achieve" (2008 SNA; see European Commission et al (2009), pp 9–10).

Today, **the importance of micro and granular information is now fully recognised**, not just by those market supervisors tasked with monitoring individual institutions ("suptech"; see Broeders and Prenio (2018)), but also more broadly by all the public authorities in charge of macro policies in the economic and financial sphere (Box 1). A key reason is that financial stress experienced at the level of individual entities, transactions or instruments can quickly reverberate into the entire financial system and beyond. Macro-level analyses should therefore be complemented with a micro-level, or in cases granular, approach – "that is, we need to see the forest as well as the trees within it" (Borio (2013)). More precisely, the high granularity of certain data sets can help: (i) "zoom in" and get idiosyncratic information that is important from a system-wide perspective; (ii) give a better sense of the distribution of economic aggregates; (iii) assemble detailed information bricks to compile more precise "macro" estimates; (iv) facilitate the design and evaluation of evidence-based policies, especially as regards financial sector reforms; and (v) trigger new, innovative ways of looking at economic and financial phenomena (see Tissot (2018)).

Box 1

Granular and micro data

Micro data and granular information relate to two close but different concepts – an important distinction as the kinds of analyses will differ depending on the type of layer (micro or granular) they are run on.

As regards ***micro data***, the traditional approach is to consider that it is "an observation data collected on an individual object - statistical unit" (OECD Glossary of Statistical Terms).[1] In the area of economics and finance, this will comprise "data on individual reporting units or specific transactions/instruments, which in most cases allow the identification of individual entities and are therefore considered confidential. In addition, publicly available data on individual reporting units are considered non-confidential although they can still be subject to data sharing limitations due to commercial property rights" (IAG (2017)).

Financial micro data will, for instance, typically refer to (confidential or not) institution-level information such as balance sheets or P&L accounts which reflect the activity reported by individual players like banks or non-financial corporations – covering, for instance, information from central balance sheet data offices (IFC (2017b)).

Granular data can be defined as comprising "disaggregated data and micro data" (IAG (2017)). ***Disaggregated data*** are "data below the level of aggregated data[2] and with a higher likelihood of identifying individual reporting units than in the aggregated data". So granular data cover less aggregated data than traditional statistics plus micro data, including when these are anonymised for dissemination.

In the financial sphere, the concept of granularity will, for instance, typically refer to data on individual transactions, eg on loans or securities, that are an input to the production of aggregated information by compiling institutions, and can also be an area for analyses and research on their own.

[1] ***Macro data*** being "an observation data gained by a purposeful aggregation of statistical micro data conforming to statistical methodology" according to this glossary. [2] With ***aggregated data*** defined as "data aggregates that have a low likelihood of identification of individual reporting units, such as those found in traditional datasets" (IAG (2017)).

After more than a decade since the GFC, **what have been the results of the various and ambitious initiatives undertaken to collect granular information on the financial system?** Central banks and financial supervisors have been at the forefront of this journey, for instance to ensure greater consistency between new micro level data sets and traditional macro aggregates, adapt statistical frameworks to the rapidly evolving financial system, and exploit micro, firm-level data sets for financial stability work.[4] For instance, many jurisdictions around the globe now have at their disposal very large and granular loan-by-loan databases, representing the bulk of what is considered "big data" by central banks (Schubert (2016); IFC (2017a)). Information on derivatives transactions reported by trade repositories (TRs) is another example of the detailed data sets that are increasingly of interest to central banks (IFC (2018)).

Yet the task of integrating (granular and) micro financial information in macro frameworks has proved more complex than initially thought. It remains hindered by the limited availability of reliable and timely reports at the level of individual institutions, especially in the least regulated corners of the financial system. Moreover, the lack of international harmonisation continues to challenge the collection of comparable, entity-by-entity data among corporations, not least across jurisdictions. Furthermore, the underlying financial statistical infrastructure is still incomplete, reflecting the slow development of global identifiers, standards for exchanging information, and data sharing arrangements – despite notable progress since the GFC as regards the Legal Entity Identifier (LEI; see LEIROC (2016)), the Statistical Data and Metadata eXchange standard (SDMX; see IFC (2016a)), and the actual international sharing of granular information on global financial institutions (see Bese Goksu and Tissot (2018)).

The difficulties above were particularly evident when Covid-19 struck, for instance during the related turmoil observed in financial markets at the beginning of 2020 (FSB (2020)). The pandemic also highlighted that a wealth of available micro data might not be fully exploited. Yet the response of official statisticians in various countries showed that "quick wins" could be achieved by making better use of the information already collected (Ducharme et al (2020)); for instance, to complement conventional survey statistics that became disrupted, to provide detailed statistics on target populations with richer insights on their distribution, or to shed light on the unexpected consequences of the pandemic (eg assessment of mobility patterns and consumer habits; De Beer and Tissot (2020)).

Moreover, **the growing interest in using granular information from alternative sources has to be balanced with the fundamental principles that govern the production of appropriate and reliable official statistics** and adhere to certain professional and scientific standards (United Nations (2013)). For instance, a key objective should be that the micro data involved are unbiased and produced in a transparent way with adequate metadata information. Another is to ensure that the provision of this kind of information is actually made independent of the data

[4] See IAG (2017) for a description of relevant experience by countries as well as regional and international organisations that illustrates the efforts undertaken in the various related areas covering the use of common statistical identifiers; the exchange of experience on statistical work with granular data and improvements in transparency; the balancing of confidentiality and user needs; the linking of different data sets; the provision of data at the international level; the ways for improving data-sharing of granular data; and the collection of data only once.

providers' business goals. A third aspect is to favour benchmarking exercises, especially against other types of data, to ensure accuracy and appropriate context. A final point is to certify that the integrity and confidentiality of the data is respected and that the information is used honestly.[5]

The above considerations clearly underscore the importance of improving the integration of micro data into macro frameworks. A relevant initiative from this perspective was the joint organisation by the Central Bank of Malaysia, the ECB and the IFC of a satellite seminar on "Post-crisis data landscape: micro data for the macro world", held on the occasion of the 62nd biennial World Statistics Congress (WSC) of the International Statistical Institute (ISI) in Malaysia in 2019. Participants from about 50 organisations, including central banks, national statistical institutes/offices, international organisations, the private sector and academia, convened to jointly analyse this topic. The event proved to be another important milestone in the IFC's ongoing work on micro-level information – especially in finding appropriate sources, developing new methodological concepts and techniques, compiling policy-relevant indicators and making use of them, and taking advantage of rapid improvements in technology (the "big data revolution") (IFC (2016a,b,c)). It also provided another opportunity to highlight existing best practices and potential opportunities, especially to support policymaking, as well as to take stock of the challenges to be addressed as a priority.

In particular, and as emphasised by the senior panel invited to conclude the seminar, **there are clear limitations in analysing economies based solely on the supply of macro statistics**. In contrast, a growing and lasting demand for more granular insights has emerged from a wide range of users – for instance to monitor risks in the financial and non-financial sectors, analyse interconnectedness and assess international spillovers.

Addressing this supply/demand information gap puts a premium on finding effective **ways to drill down from aggregates to more micro information**. This can offer a wide range of analytical opportunities, by: (i) reducing the risk of overseeing phenomena that may not be noticed at the macro level; (ii) better supporting the design of macro policies, especially in the areas of monetary and financial stability; and (iii) allowing for more accurate measurement of the impact of authorities' actions. In turn, such monitoring would also enable some fine-tuning where and when appropriate. Moreover, making better use of granular (including micro) data collections should be favourable for reporting agents, by lowering their reporting burden with more stable and resilient requirements and reducing the need for them to compile aggregated data. However, harmonising and streamlining data reporting is essential to reap those benefits, and corresponds in fact to a strong demand expressed by the financial industry.

Looking ahead, to make concrete progress attention should focus on four main tasks: building up new micro data collection frameworks (Section 1); accessing and making use of more granular sources of information (Section 2); developing new and adequate analytical tools (Section 3); and bridging the gap between micro- and macro-level statistics (Section 4).

[5] See eg the Research Mission Statement and related Principles published by the US online real estate company Zillow, whose data are now being used by the Board of Governors of the US Federal Reserve System for official statistical purposes; see Hume McIntosh (2021).

1. New frameworks for collecting micro data

A first, obvious issue for producers of official statistics is how to organise themselves to collect the increasing amounts of granular information which are potentially available. This is particularly important for central banks and supervisory authorities, which are confronted, on the one hand, with increased reporting fatigue among respondents and, on the other hand, with a steady rise in data needs to conduct evidence-based macro policies and monitor the financial sector.

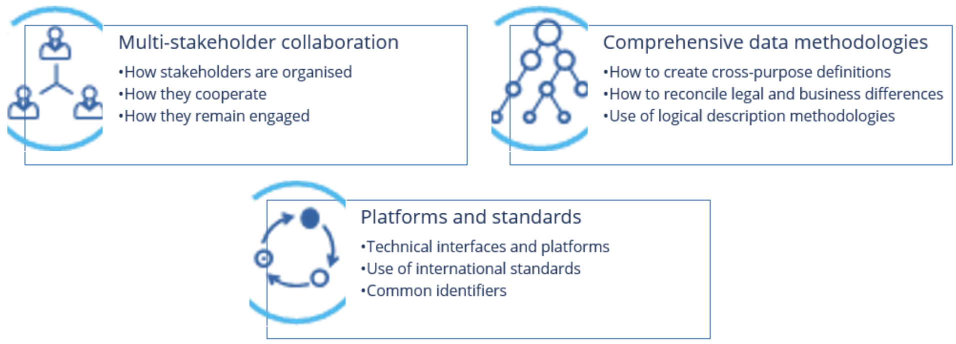
Various countries' experiences suggest that this trade-off can be best addressed by **integrating reporting requirements and standardising data in generic "cubes"**, ie structures that allow for the storing, reporting and analysing of data points with multiple dimensions, so as to cover the wide range of reporting exercises and analytical exercises. Such an approach can also allow new information needs to be addressed with more agility and less effort as they come in.

According to Eurofiling, a forum initiative started in 2005 to improve collaboration on and awareness of European regulatory reporting among regulators, supervisors and entities from both the public and private sectors, **a key lesson of the qualitative case studies for recent international projects is that there are three main factors necessary for success** (Chart 1). First is the need to have effective collaboration between the multiple parties involved. To ensure high-quality inputs in data collections, central bank statisticians should in particular understand how the various reporters are organised, how they can cooperate, what should be done so that they remain engaged over time, and how data platforms should be set up to support the provision of continuous feedback. The second main issue is the demand for comprehensive data methodologies and models. Since statistical compilation chains involve many stakeholders, it is important that methodological definitions are well documented, compatible and applied consistently for the variety of purposes considered and with sufficient agility (for instance to reconcile the legal and business approaches supporting data collections). A third issue is to develop adequate platforms and standards for data compilation. This calls for the development of comprehensive technical interfaces (for instance between firms' internal business data warehouses and their submissions for official statistics reporting), shared standards (ideally at the international level since the financial system is global) and common identifiers to correctly identify financial agents as well as their transactions and products.

Certainly, **a number of important initiatives are already being developed in several central banks to address these issues**. In terms of methodologies, the Bank of Spain has built up a Credit Registry Dictionary and has worked on using a data point model, based on the XBRL (eXtensible Business Reporting Language) standard, to streamline its data collection. Singapore has developed an initiative to foster a common understanding of supervisory and statistical reporting requirements in the financial industry. Turning to Europe, the ECB has designed a single data dictionary, the Banks' Integrated Reporting Dictionary (BIRD), to foster cooperation with the banking industry in the field of regulatory reporting, alleviate the reporting burden and improve data quality. This has been complemented with the provision of a large set of precise definitions that are operational and can even help beyond regulatory reporting – for instance, to support the compilation of the Analytical Credit Database (AnaCredit)) (ECB (2019)) – and the use of a single multidimensional data model (SMCube) based on standard identifiers such as the LEI and ISIN (International

Securities Identification Number), in turn allowing the same micro data to be used for multiple purposes. Similar efforts have been conducted by the European Securities and Markets Authority (ESMA), which is the rule setter for securities and derivatives markets in Europe, for instance to design a reporting framework for derivatives transactions through TRs combining methodological definitions, the description of the related reporting obligations and the use of an ISO standard. On its side, the European Commission has also been working on enhancing financial data standardisation and has also developed a fully fledged "European strategy for data", especially as regards the financial data space (see European Commission (2020), pp 30–1).

Three main paradigms for modern statistical data frameworks: a view from Eurofiling Foundation — Chart 1

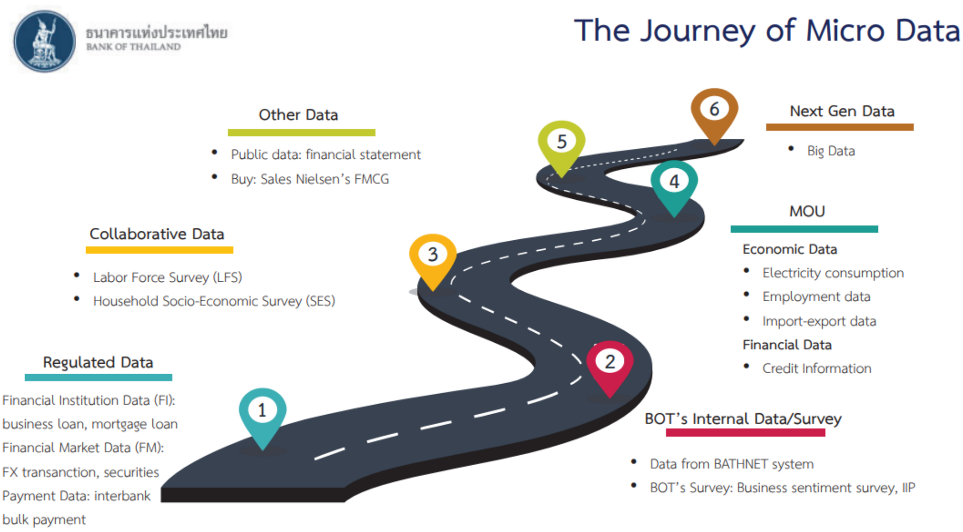


Source: M Piechocki, "Paradigms for modern statistical data frameworks and systems", *IFC Bulletin*, no 53, April 2001.

In addition to those strategies aiming at streamlining data reporting processes, **it may be useful to define a comprehensive data strategy for the institution in charge of data collection**. This should encompass the related business cases, the projects envisaged (eg building up a data platform), as well as the related governance issues faced, including data sharing/access and collaboration exercises. The experience of the Bank of Thailand is that such a strategy can bring many benefits. The reason is that information is a strategic asset and that organisations could make more active use of it. The central bank has therefore developed a multi-year plan with the goal of improving the stocktaking, analysis and sharing of the various economic and financial micro data potentially available in the country – especially through greater use of high-frequency data from various sources including unstructured ones, eg text (Chart 2). The primary focus has been on micro-level statistics, since their granularity can provide various insights for policy use, for instance to analyse monetary policy mechanisms or developments in property markets, as well as to improve the efficiency of the market operations conducted by the central bank. Certainly, this project has faced many challenges, especially regarding staff skills (to be developed through external recruitment or internal training), the selection of technology to handle vast and various types of data sets (eg structured as well as unstructured), the development of adequate legal and data governance frameworks, and the need to comply with data protection rules.

Taking stock of micro data sources: the experience of the Bank of Thailand

Chart 2



Source: P Chansang, "Data management in the data evolution era at Bank of Thailand", *IFC Bulletin*, no 53, April 2021.

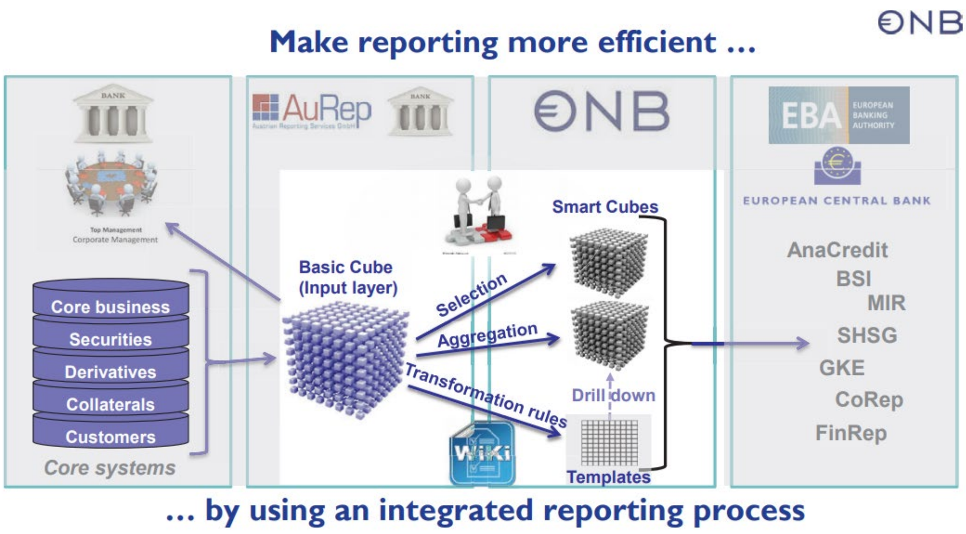
Turning to the other side of data collection processes, reporting institutions are also making active progress. This is particularly the case for those banks and other financial institutions that have faced increasing reporting requirements, especially in the wake of the GFC. The resulting burden has been challenging their traditional way of collecting data from internal reports, which were typically organised along separated business processes – with the risk of redundancies, inflexibilities and inconsistencies between different reporting exercises – and were rarely standardised (implying that any new collection exercise would lead to extra reporting costs).

The experience in Austria is that one way to overcome these issues and make data collection exercises more efficient on the reporting side is to better integrate the reporting process (Chart 3). To do so, **the central bank has developed a "regtech approach"** (which basically refers to the provision of methodology, technology and processes to financial institutions to support regulatory monitoring, reporting and compliance). An important factor in this approach was the decision by the largest Austrian banks – in coordination with the central bank – to establish a central reporting platform (AUREP). The goal was to ensure that all the data to be reported by commercial banks are sourced from a general ledger – that is, a general "basic cube" (input layer) with ad hoc validation rules to support data quality management. This allows storing and organising all the information in "smart cubes" made of specific data selected/aggregated/transformed so that these are easily available for external reporting exercises with different types of aggregation level (eg consolidated, sectoral). This setup can also be used effectively to support the banks' internal reporting.

Such an approach can help to improve data quality, achieve synergies and lower reporting costs, and can **represent a "game-changer" with an overall impact on the information organisation framework**. In addition, it facilitates the use of advanced analytical techniques, as data can be accessed and retrieved more easily. Lastly, one key lesson is that the more granular the reporting requirements, the more effective the approach can be, since the underlying data are not transformed when stored and can thus be re-used for different exercises – minimising the potential impact of additional enquiries by the authorities. Further progress could be obtained if regulators set out their reporting requirements in standardised languages, possibly machine-readable ones. This would allow retrieval (or "pull") of the data from the reporters' own internal cubes, instead of asking them to send (or "push") the data themselves – in turn reducing reporting costs and further increasing data quality and timeliness.

Integrating the data reporting process: the experience of the Central Bank of the Republic of Austria

Chart 3

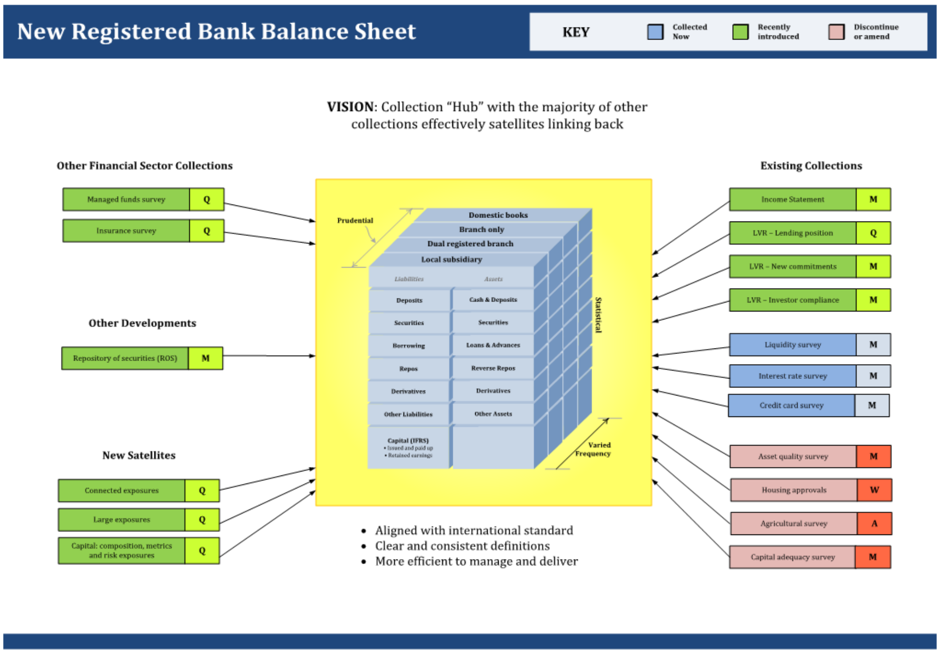


Source: J Turner, "OeNB's reporting data model as RegTech/SupTech solution", *IFC Bulletin*, no 53, April 2021.

The focus on reducing reporting burden is also high on the agenda of the Reserve Bank of New Zealand, which has developed a "collect once, use multiple times" approach similar to the one adopted in Austria (Chart 4). A main goal has been to enhance the accuracy of the data collected from the financial industry, especially banks, and ensure that all regular data collections are relevant, fit for purpose and cost-effective. In particular, and to avoid misunderstandings in implementing the related instructions, the Reserve Bank ensures that reporting exercises are designed in strong connection with the data providers. Building relationships, providing clarity around data requests and spending time discussing with data providers is not costless but is proving extremely valuable in the long run. Moreover, any new specific data request – covering all prudential, macroprudential and statistical purposes – can be linked back to a single high-level balance sheet or central "hub". The experience is that this "collect once but use it multiple times" approach can be instrumental in reducing errors in the data sent, addressing new reporting needs with more agility, and facilitating data analytics work.

Single collection for multiple use (prudential, macroprudential and statistical purposes): the experience of the Reserve Bank of New Zealand[1]

Chart 4



[1] As at 2017.

Source: S Schuster and N Humphries, "Collect once, use multiple times – the Reserve Bank of New Zealand's data collection approach", *IFC Bulletin*, no 53, April 2021.

2. Accessing and using micro data sets

Once the (micro) data sets are collected, the next steps are to enable easy and effective access of this information (overcoming the challenges related to size and complexity) and to make sense of it. In other words, it is not sufficient to "collect" the dots, what is essential is to "connect" them (Caruana (2017)) – basically meaning central banks should perform relevant analyses that can support both the preparation and monitoring of their policies. These include monetary policy, typically at the heart of central banks' missions, but also increasingly macro- (since 2011 for the ECB) as well as microprudential policies (since 2014 for the ECB in the area of banking supervision). The need for adequate information supporting these three different functions has clearly reinforced the interest in collecting granular data. As stated by the former President of the ECB, well established macro statistics will continue to provide the "big picture" but "we should also offer a magnifying glass" (Draghi (2016)).

However, the information provided by very rich granular data sets can be overwhelming. Making it usable requires the support of ad hoc techniques so as to transform simple "data points" into "knowledge" (Drozdova (2017)). A key success factor is therefore to have adequate staff skills (eg data scientists) as well as appropriate IT tools (eg big data analytics; Wibisono et al (2019)). Another important lesson from central banks' work with micro data is the importance of combining different approaches, requiring the setup of multidisciplinary teams involving IT staff, subject matter statisticians and economists. One reason is the large number of different types of micro data sets that require specific competencies to be properly accessed and analysed; for instance, commercial data, administrative data and social media data sets. In addition, the interest of central banks' users has also expanded, so that projects dealing with granular data have become more frequent and serve increasingly multiple purposes – from the traditional research area to using micro data as an auxiliary input (or even sometimes as the core input) for the fine-tuning of macro policies and micro surveillance processes. Furthermore, statistical dissemination has become more complex, as an impressive and increasing wealth of information is made available to internal users and the research community. This has important consequences for the organisation of central banks' data work. One such consequence is that information architectures need to allow for combining multiple data sets, sharing data throughout the institution, and supporting various tasks such as regular statistical compilation, ad hoc data exploration and experimental activities. Another is that close cooperation is required among the various stakeholders involved, so that different sources of information can be used in a complementary way and/or for different purposes.

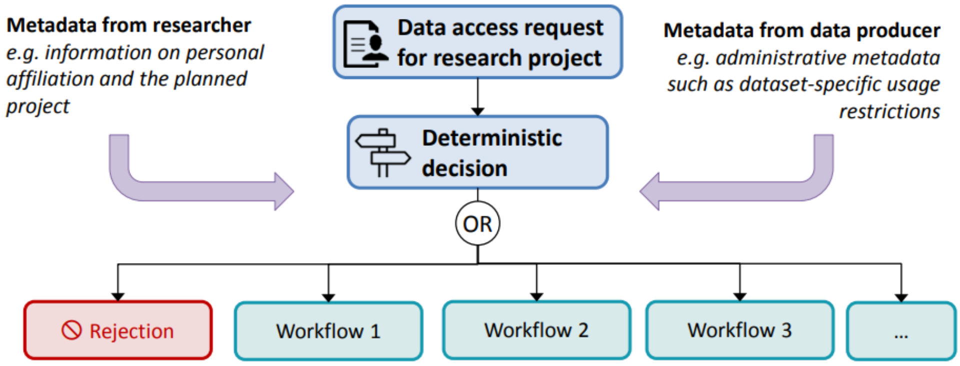
These issues can pose important challenges in practice, not least because of the resource constraints faced by public institutions. Hence, while the use of aggregated data sets has become routine for those authorities in charge of monitoring macroeconomic developments and implementing related policies, progress in extracting insightful developments from micro data sets has been much more limited. In particular, while one can relatively easily conduct standard macro aggregations out of individual data appoints, the identification of patterns of interest out of heterogeneous developments combining large granular data sets can be a much more complex task.

One solution to these challenges is knowledge-sharing. To facilitate that, several central banks, together with a number of international organisations and statistical offices, have formed the International Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA), supported by the BIS (Bender et al (2018)). A main objective of this initiative was to identify best practices in facilitating access to, and sharing of, micro data sources, for instance to develop data access procedures for researchers, as well as to explore new analytical possibilities, eg to analyse the effect of policies. This review has highlighted the benefits of automating decisions in workflows, by: (i) relying on effective "metadata schema" that describe how administrative data sets can be managed and used from a legal, organisational or technical point of view; and (ii) having workflows that follow such metadata information deterministically (Chart 5). Lastly, the international statistical community, including central banks, is also actively engaged in public sharing of IT tools in the context of SDMX (IFC (2016a)) as well as software codes through secure online software repositories (eg based on the free, open source system Git). This can greatly support the re-use of codes in open source languages such as Python and R for accessing and exploring micro data sets without requiring the actual exchange of the underlying data.

Yet a key lesson is that the devil is often in the details: the various pilots implemented by central banks have faced **concrete implementation challenges** – including the lack of standardisation of the information contained in the data sets and of the related workflows ("metadata reference"), the difficulties in setting up adequate facilities allowing external users to access this information, and the management of confidentiality risks when results drawn from the work on micro data have to be disseminated.

Towards automation of decisions in workflows: INEXDA best practice recommendation

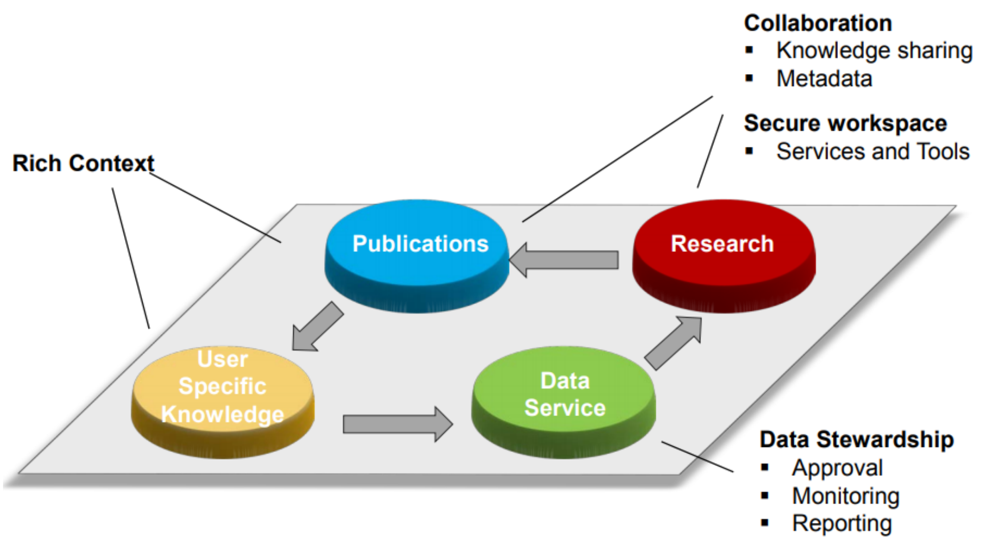
Chart 5



Source: S Bender, "Best practice recommendations on facilitating access to microdata: outcomes from the INEXDA working group on data access", *IFC Bulletin*, no 53, April 2021.

In fact, a number of successful national initiatives are worth highlighting. One relates to **dealing with external research projects using micro data, an area in which the Deutsche Bundesbank has developed significant expertise in recent years**. A key lesson is the importance of enhancing user knowledge of the data sets in question by providing rich contextual information (or "metadata"). This can be achieved through multiple steps forming a feedback loop, replacing previous one-directional information flows (Chart 6). First, by collecting information on the usage of the respective data sets; for instance, to know who has been using a specific type of data in the past and for what purpose. This can be achieved by, first, moving from "classical" to "data-centric metadata" that includes information on the use of the data and on the related administrative workflows. Second, by learning from this experience in an automated way, eg by relying on ad hoc ML techniques; for instance, to inform a user working on a specific topic what data sets are also of interest for other researchers conducting similar work – providing insights such as "people working in your area are most frequently interested in these fields". And third, by actively engaging with data users to incorporate their specific knowledge and findings and, in turn, improve the data sets' metadata information, so as to facilitate future work – eg information related to the data viewed by the user, similar data of interest, what others have done with similar data ("recipes") and what alternative recipes may be.

The knowledge life cycle enabling feedback loops | Chart 6



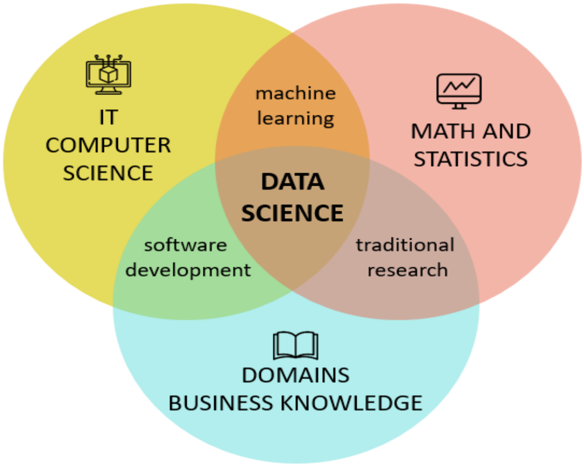
Source: C Hirsch, "Enhancing user knowledge by providing (richer) context to microdata", *IFC Bulletin*, no 53, April 2021.

Another interesting initiative has been to facilitate the working of teams with different competencies to make the most of the new data landscape. The objective is to have specialised staff, represented by different "colours", joining forces and leading to the assembling of mixed business and technology skills in composite groups of "purple people", ie those specialised staff whose skills overlap. To ensure such an adequate blending of staff skills, the Banco de Portugal has developed a tailored internal training programme in partnership with a university, noting that attracting external specialised skills like data scientists had proved to be an unsuccessful task (Chart 7). Similarly, the Netherlands Bank has developed a "Data and Technology" traineeship with the goal of developing "datapreneurs".

Another important point is **collaboration**. The new Integrated Data Management initiative – undertaken by Banco de Portugal as part of its strategic planning aims – promotes the creation of added value by making users, statisticians, data scientists, IT staff, supervisors etc work together. This collective approach is reported to have enhanced the use of micro data in the institution, removed cultural/organisational obstacles, and rationalised the various processes involved in terms of data collection/processing/sharing. It was also a key element supporting the Bank's internal governance model, with data stewards allocated to all business areas and tasked with overseeing the quality and fitness for purpose of the data assets. Turning to the Netherlands Bank, the focus has been on promoting discussions between data scientists and supervisors and policy officers, with the organisation of joint projects in a "hub and spoke" model.

The Banco de Portugal School of Data Science: the internal training programme to provide business areas with a blend of staff skills

Chart 7

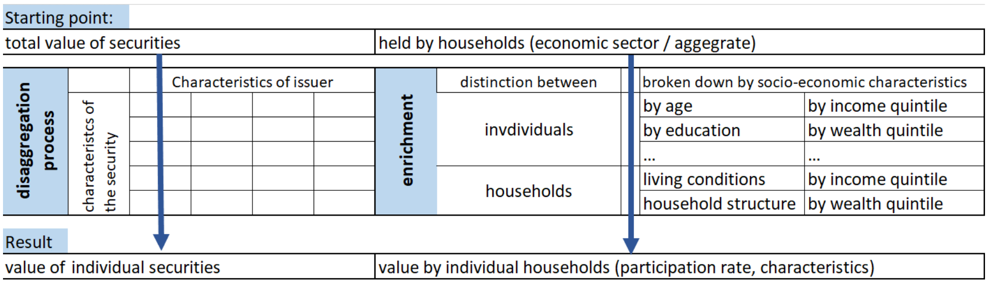


Source: L Teles Dias, "Post-crisis skills landscape: The emergence of 'Purple People'", *IFC Bulletin*, no 53, April 2021.

The good news is that these efforts can rapidly pay off, with central banks increasingly able to develop informative use cases based on the micro data sets stored in their vaults (La Cava (2015). Moreover, experience shows that this information can improve understanding of the economy's functioning – eg to facilitate macroeconomic forecasting exercises that are based on the modelling of economic agents' behaviour (Holmquist and Hume McIntosh (2020) for a US illustration). In Austria, for instance, efforts to "bridge" micro results with developments in macro aggregates have supported the analysis of the financial behaviour of households. While macro statistics based on financial accounts had been able to document significant trends in recent years (the "what"), such as observed shifts in deposits' maturity and financial instruments, more granularity was required to understand the factors at play behind these phenomena (the "why"). In contrast, the micro level information drawn from the Consumption and Financing Survey can help identify the types of households behind observed aggregated portfolio changes (Chart 8). In particular, this approach highlighted that the shift from Austrian deposits with longer maturity to overnight deposits could be observed for nearly all types of households; but less than 10% of them (those with higher net wealth) were responsible for the aggregated shift from debt securities to investment fund shares. The work to enrich aggregated information for socioeconomic characteristics has also emphasised the need to ensure methodological and scope consistency between the different types of data.

Enriching macro aggregates by socioeconomic characteristics using distributional information from micro-level household survey data

Chart 8



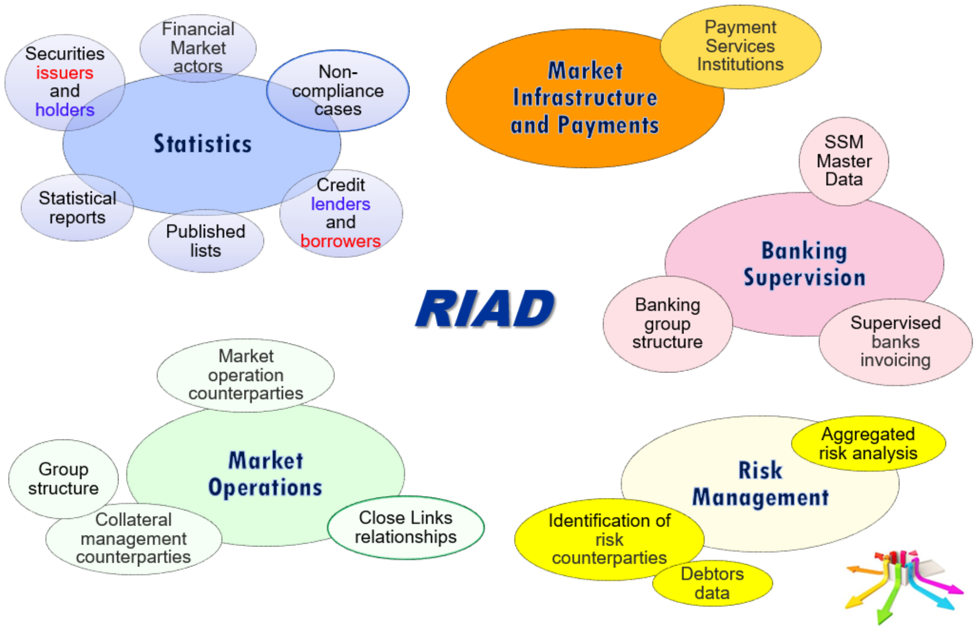
Source: J Turner, "Financial behaviour of households under the lenses of distributional aspects: the OeNB's experiences", *IFC Bulletin*, no 53, April 2021.

Another successful example relates to <u>ECB work on granular data on credit and credit risk</u>. Instead of developing new ad hoc surveys that may lead to inconsistent pictures, the idea has been to pull together all the granular information of potential interest in one data set (collected from AnaCredit) that can support different analyses and policy uses (Israël et al (2017)). For instance, such a rich data set can enable analysis of supply and demand credit factors depending on market conditions, small and medium-sized enterprises' specificities, the transmission channel of monetary policy, sectoral risks in the central bank's asset management operations etc. This wide range of analyses can thus assist various functions, eg monetary and macro- and microprudential policies as well as emerging policy needs related to environmental, social and governance factors. Moreover, the exercise provides value by allowing different data sets to be connected, with a holistic approach favouring consistency and interoperability across data domains. Furthermore, it was an opportunity to develop new / enhance existing statistics; for instance, data on credit proved to be more accurate when linked to a reliable business register.

To this end, the **Register of Institutions and Affiliates Database (RIAD)** was set up as the backbone of all granular statistics for the European System of Central Banks, with a unique master data set, a shared platform, high-quality checks and rules, and a comprehensive data model including reference data on individual units and relationships among them (Chart 9). However, successfully combining different data sets requires clear data definitions and methodologically sound reference information (metadata), in turn highlighting the importance of establishing a good dialogue with the financial industry (see Section 1).

Standardising and integrating existing frameworks across domains: the ECB RIAD register to integrate micro and granular data

Chart 9



Source: J-M Israël, "Linking micro datasets to better service policy-making and analyses", *IFC Bulletin*, no 53, April 2021.

3. New analytics for working with micro data

Due to their inherent complexity and size, granular data sets are difficult to work with, as they can require substantial time and IT capacity. This is indeed why the main applications on micro data pursued in the past have basically related to the extraction of aggregated indicators, typically using simple Structured Query Language (SQL)-based queries designed for "traditional" relational systems. **Yet increased computing power capacity and new analytical tools such as ML techniques have clearly changed the picture**, as it has become easier to work directly on the underlying data points to perform various purposes, from pure statistical production tasks to more in-depth and/or real-time complex analyses. In turn, the insights gained in understanding, for instance, the functioning of complex financial systems, assessing potential vulnerabilities and analysing contagion effects can help to effectively prepare and monitor policy actions.

However, several challenges have to be addressed. First is data access: one has to identify adequate sources, set up appropriate sharing arrangements, and implement data governance procedures, not least as regards the ownership of, and access to, sensitive information. A second issue is whether to opt for on-premises or cloud-based solutions. The latter are increasingly popular, not least because of their scalability and flexibility. In addition, while they can entail security risks, most notably leakages of sensitive data that may have high reputational costs for central banks (IFC (2020)), the recent experience with the Data Science Hub set up by the

Netherlands Bank shows that ways can be found to save confidential data in the cloud. A third issue is the "black box" syndrome: one has to have a good understanding of the techniques involved in order to ensure that analyses are not only accurate but also "interpretable" – so that specific explanatory causes or factors can be identified and communicated as policy-actionable "causal stories". This again emphasises the importance of the human factor.

Despite these challenges, **one important avenue explored by central banks in making using of new analytical tools relates to their role as producer of official statistics. In particular, this is used to enhance data quality management (DQM) processes** – performed in general through both automated checks (to verify pre-determined relationships) and plausibility checks (by relying on "acceptance regions") – eg by setting up imputation techniques, detecting outliers and duplicates. Yet, and in contrast to traditional macro statistics, the checking of large granular data sets often characterised by heterogeneous reporting patterns can be cumbersome, complex and almost impossible manually.

Automated techniques can greatly facilitate these tasks, as shown by the experience of the Bank of Italy in conducting ML-based quality checks on granular banking data (Chart 10). The approach, based on a supervised learning algorithm (Quantile Regression Forests), was to detect abnormal patterns in payment services data reported by banks. The results, cross-checked with reporting agents, show that this can help detect "new outliers" (ie those which had not been initially identified by existing data quality management processes) more precisely and with reasonably high accuracy. Moreover, the setup of dynamic acceptance thresholds and periodical training of the algorithms allow for an automatic update as new data come in, reducing analysts' involvement.

A supervised learning algorithm to detect outliers in payment services data reported by banks: the potential for more precise quality checks

Chart 10

Cross check of outliers with banks

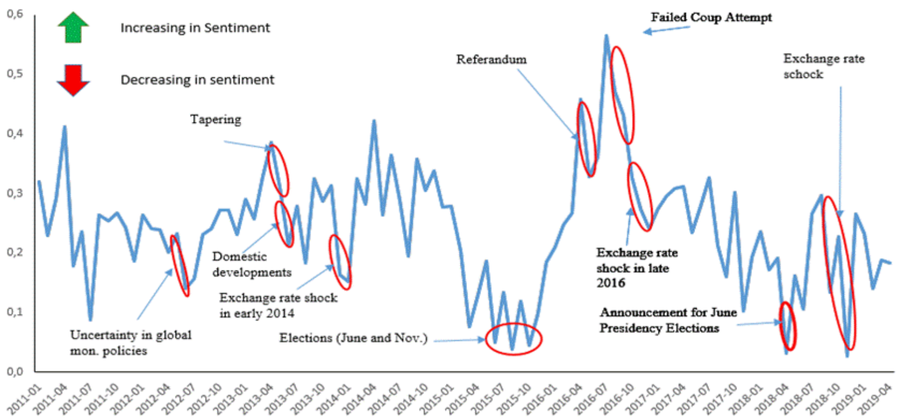
Prediction intervals:	PI_1 $[q_{0.01}, q_{0.99}]$	PI_2 $[q_{0.025}, q_{0.975}]$	PI_3 Inter-quartile range
a-Total number of potential outliers	**373**	**489**	**457**
b-Anomalies detected and revised ("true positives")	289	312	292
c-Confirmed observations ("false positives")	84	177	165
d-Precision b/a (%)	**77.5%**	**63.8%**	**63.9%**

Source: F Zambuto, "Quality checks on granular banking data: an experimental approach based on machine learning", *IFC Bulletin*, no 53, April 2021.

Another important application is to make sense of the vast amount of textual information to extract summary indicators, for instance economic sentiment. This kind of approach has been implemented at the Central Bank of the Republic of Turkey to predict economic growth based on subjective information such as emotion, opinion and attitude (Chart 11). The input information was extracted from published statements based on a lexical approach and the use of ML techniques like natural language processing (NLP) for the automated processing and analysis of large amounts of text – with the setup of a term matrix to generate a vectoral representation of the document of interest, and the subsequent extraction of a summary sentiment indicator.

Evolution of the sentiment index for the Turkish economy

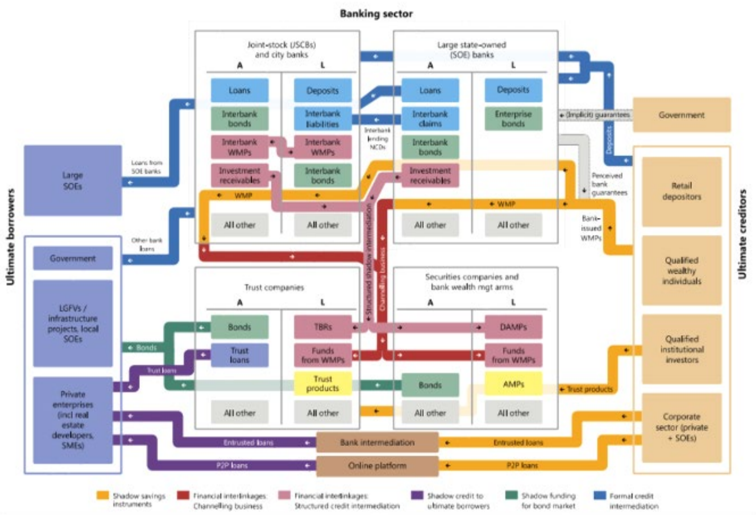
Chart 11



Source: B Akisoglu, "Predicting economic growth using machine learning techniques and sentiment analysis", *IFC Bulletin*, no 53, April 2021.

A third important area relates to the analysis of financial networks with the identification of their topological characteristics, such as the likelihood of connection between specific nodes ("connectivity") or neighbours ("clustering"), the distribution of the shortest paths among nodes ("betweenness centrality"), and the relative importance of links and nodes in the system ("average degree"). Micro-level information can be particularly suited for forming a comprehensive map of an ecosystem and its interconnections, for instance to analyse the financial system of China (Chart 12). Moreover, and as experimented by the Data Science Hub set up by the Netherlands Bank, it is an important building block for conducting stress test exercises. These typically involve designing severe circumstances that have yet to materialise, and the flexible combination of granular data from different sources to perform the assessment of such hypothetical scenarios. For instance, individual banks' data on stocks (balance sheets) and flows (transactions) can help identify troubled firms, sectors or markets, allowing the propagation of potential shocks to be traced through the financial sector.

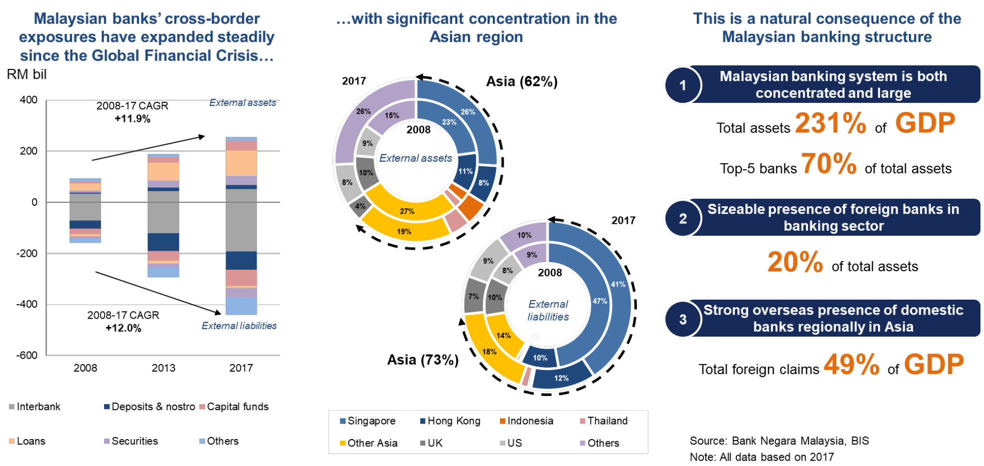
China shadow banking map: stylised structure of claims Chart 12



Source: T Ehlers, S Kong and F Zhu, "Mapping shadow banking in China: structure and dynamics", *BIS Working Papers*, no 701, February 2018.

The Central Bank of Malaysia has followed a similar approach, with **a more global perspective so as to analyse international linkages and cross-border contagion risk** given the growing regionalisation of domestic banks and the significant presence of foreign banks in Malaysia (Chart 13). The work was based on granular data provided by banks which were used to develop stress-tests to: (i) assess banks' vulnerability to external shocks at the macro and micro level; (ii) estimate contagion paths across banks (that is, how the shocks propagate themselves); and (iii) identify potential systemic linkages and cross-border domino effects between internationally active banks and Malaysian banks. The results show that the solvency contagion risk of Malaysian banks to external shocks has declined in recent years, but that potential sources of contagion remain. The approach also suggests that a micro-based network analysis approach, combined with counterfactual simulation, could usefully complement macroprudential surveillance tools and support the development of effective arrangements for crisis containment, management and resolution in case unexpected shocks happen. Lastly, the modelling highlights the benefits of cooperation in terms of information-sharing and surveillance to effectively address cross-border contagion risks.

Assessing cross-border contagion risk for Malaysian banks Chart 13



Source: H Sababathy and L Sheng Ling, "Cross-border contagion risk to the Malaysian banking system", *IFC Bulletin*, no 53, April 2021.

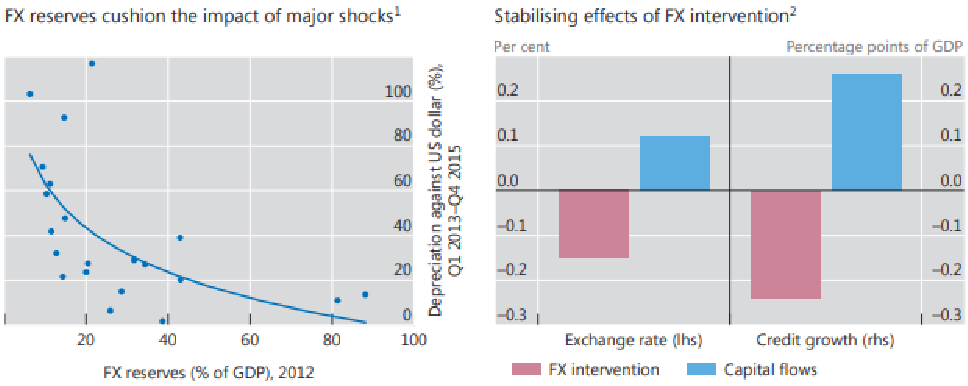
4. Bridging the gap between macro- and micro-level statistics

There has been a growing interest among central banks and financial supervisors in directly conducting analyses in a very granular way, especially to better prepare the actions of public authorities and help measure their effectiveness, as documented above. **Yet a key point of interest for policymakers is how to link developments on the individual institution or household level to the more general picture**. Granular data sets can surely not be a substitute for aggregated statistics, but their input can add value for analysing macroeconomic issues, for instance the functioning of the complex financial system, the evolution of inequalities, or sectoral issues.

In order to better understand the financial system, bridging the gap between micro- and macro-level statistics can be instrumental for assessing financial stability risks, a key issue for central banks (Crockett (2000)). The reason is that a financial crisis is by definition a system-wide, macro event that can be caused by very specific, micro factors, for instance at the level of financial institutions that have systemic importance. The challenge is therefore to extract signals that can have more general implications from the wealth of information on idiosyncratic firms. Research conducted at the BIS on predicting and preventing financial crises suggests that a good starting point is to assess the risk of individual bank distress by using bank loan data from credit registries to identify credit demand and supply factors. From there, one can draw useful insights on the general effects of policies, including on their respective contributions (Chart 14). For instance, the use of countercyclical macroprudential tools may address financial imbalances in a targeted way, while monetary policy instruments can affect the economy in a more indistinct way because of their general impact on credit or debt service ratios. Moreover, the granularity of the data can also be helpful in qualifying the results observed at the global level, eg to assess the special challenges faced by emerging market economies. One example relates to the Latin American region, where the effectiveness of financial stability policy actions is reported to be stronger when monetary and macroprudential policies are synchronised.

Using micro data to facilitate the understanding of the effects of macroprudential and monetary policies

Chart 14



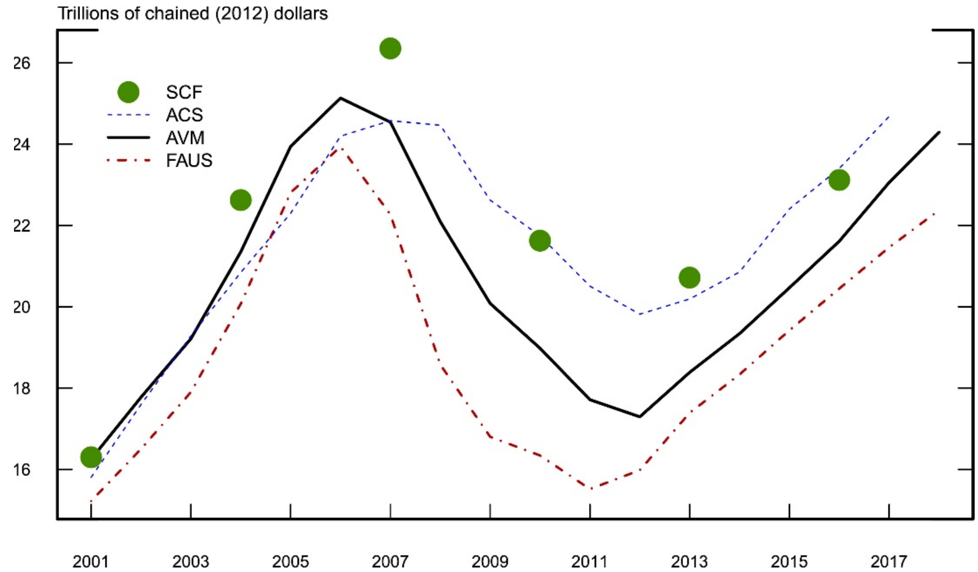
[1] Based on 21 EMEs. [2] Coefficients from a panel regression analysis for 20 EMEs from 2000 to 2017. The dependent variable is the percentage change in the bilateral exchange rate against the US dollar (increase denotes an appreciation) and the change in the ratio of domestic credit to GDP, respectively. The regressors are the accumulation of FX reserves as well as net capital inflows, respectively, as a ratio to GDP. The control variables comprise the lagged dependent variable, the short-term interest rate spread against the United States, the log change in the VIX, the log change in the CRB commodity price index and country fixed effects.

Source: BIS, "Monetary policy frameworks in EMEs: inflation targeting, the exchange rate and financial stability", *Annual Economic Report,* 2019, Chapter II.

Turning to distributional issues, the current macro framework presents notable shortcomings. Certainly, the 2008 SNA clearly states the importance of considering the skewed distribution of income and wealth across households. But it recognises that getting this information is "not straightforward and not a standard part of the SNA" (2008 SNA, no 24.69) and that "there would be considerable analytical advantages in having micro databases that are fully compatible with the corresponding macroeconomic accounts" (2008 SNA, no 1.59). One illustration of this point is the work conducted by the Board of Governors of the US Federal Reserve System to develop Distributional Financial Accounts (DFAs), by using micro data from the Survey of Consumer Finances to distribute the aggregate household balance sheet information available from the Financial Accounts (Batty et al (2019)). Another initiative to complement macro aggregates with micro-level information has been to enhance the measurement and geographical distribution of aggregate US housing wealth, a major component of total household wealth and also an important variable to consider in quantitative macroeconomic models and analyses (Chart 15). Granular "big data" information on housing values generated by ML-based valuation models from Zillow, a private real estate and data analytics firm, was aggregated with additional data from the US Census to estimate and make public long monthly series of aggregate housing wealth (Gallin et al (2021)). Moreover, additional detailed geographical breakdowns were also publicly released. The method has proved to be robust and accurate, providing much insight into the specificities observed at local level and in a timely manner.

Measuring US aggregate housing wealth based on various micro and macro data sources[1]

Chart 15



[1] ACS: American Community Survey (US Census Bureau); FAUS: Financial Accounts of the United States; SCF; Survey of Consumer Finance (triennial); AVM: Automated Valuation Model of the Board of Governors of the Federal Reserve System based on big data collected by Zillow.

Source: Gallin et al (2021).

A final issue of interest is the better assessment of developments that are neither at the micro nor the macro level, for instance by economic sector. In Europe, work has been conducted at Eurostat to improve understanding of the factors driving sectoral productivity developments, based on a similar approach developed by the University of Groningen (Netherlands). Taking below macro-level information (Chart 16) into account has helped produce estimates of labour productivity, supporting a wide range of analyses. A key issue was the adjustment of sectoral measures for purchasing power parities (PPPs), which are indicators of price level differences across regions or countries – knowing that, traditionally, PPPs are not estimated for sectors but from the expenditure side of GDP. The International Comparison Program was set up to calculate such price level indices for a wide range of goods and services, to be used as inputs for various studies such as real GDP comparisons or the measurement of poverty rates.

Mobilising granular data sources to support industry-specific productivity comparisons

Chart 16

Prices:

- Prodcom: unit values for over 3000 manufactured goods
- PPP Programme: consumer goods and services, equipment, construction, education, health, collective services
- Agriculture prices and price indices: producer prices for 128 products
- "Proxy PPPs"

Weights:

- Prodcom: production values at product level (below BH level)
- Structural business statistics: GVA for 4-digit NACE branches (above BH level)
- National accounts: GVA for 64 NA branches (further aggregation)

Source: L Olislager, "Key indicators for sectoral productivity analysis", *IFC Bulletin*, no 53, April 2021.

References

Batty, M, J Briggs, K Pence, P Smith and A Volz (2019): "The distributional financial accounts", *FEDS Notes*, Board of Governors of the Federal Reserve System, 30 August.

Bender, S, C Hirsch, R Kirchner, O Bover, M Ortega, G D'Alessio, L Teles Dias, P Guimarães, R, Lacroix, M Lyon, E Witt (2019): "INEXDA – the Granular Data Network", *IFC Working Paper*, 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, April.

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

Broeders, D and J Prenio (2018): "Innovative technology in financial supervision (suptech) – the experience of early users", *FSI Insights on policy implementation*, no 9, July.

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

Crockett, A (2000): "Marrying the micro- and macroprudential dimensions of financial stability", speech to the 11th International Conference of Banking Supervisors, 21 September, Basel.

De Beer, B and B Tissot (2020): "Implications of Covid-19 for official statistics: a central banking perspective", *IFC Working Papers*, no 20, November.

Draghi, M (2016): "Central bank statistics: moving beyond the aggregates", speech at the Eighth Statistics Conference, Frankfurt am Main, 6 July.

Drozdova, A (2017): "Modern informational technologies for data analysis: from business analytics to data visualization", *IFC Bulletin*, no 43, March.

Ducharme, L M, J Tebrake and Z Zhan (2020): "Keeping economic data flowing during COVID-19", *IMF Blog*, 26 May.

European Central Bank (ECB) (2019): *AnaCredit Reporting Manual Part I – General Methodology*, second edition, May.

European Commission (2020): "A European strategy for data", *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*, COM (2020) 66 final.

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

Financial Stability Board (2020): *Holistic review of the March market turmoil and COVID-19 pandemic: financial stability impact and policy responses*, November.

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.*

Gallin, J, R Molloy, E Nielsen, P Smith and K Sommer (2021): "Measuring aggregate housing wealth: new insights from machine learning", *Journal of Housing Economics*, vol 51(C).

Greenwood, R and D Scharfstein (2013): "The growth of finance", *Journal of Economic Perspectives*, vol 27, no 2, spring, pp 3–28.

Holmquist, E and S Hume McIntosh (2020): "Policy uses of financial accounts: the US Fed's perspective", in *IFC Bulletin*, no 51, February.

Hume McIntosh, S (2021): "Measuring aggregate housing wealth: new insights from automated valuation models", *IFC Bulletin*, no 53, April.

IFC (2016a): "Central banks' use of the SDMX standard", *IFC Report*, no 4, March.

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

——— (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.

——— (2018): "Central banks and trade repositories derivatives data", *IFC Report*, no 7, October.

——— (2020): "Computing platforms for big data analytics and artificial intelligence", *IFC Report*, no 11, April.

Inter-Agency Group on Economic and Financial Statistics (IAG) (2017): *Update on the Data Gaps Initiative and the outcome of the Workshop on Data Sharing*, March.

Israël, J-M, V Damia, R Bonci and G Watfe (2017): "The analytical credit dataset: a magnifying glass for analysing credit in the euro area", *ECB Occasional Paper Series*, no 187, April.

La Cava, G (2015): "The development of databases linking micro and macro data – an Australian perspective", in *IFC Bulletin*, no 40, November.

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*, March.

Schubert, A (2016): "AnaCredit: banking with (pretty) big data", *Central Banking Focus Report*.

Tissot (2018): "Providing comparable information to assess global financial stability risks", *Eurostat Statistical Reports*, KS-FT-18-001, January.

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

Wibisono, O, H Ari, A Widjanarti, A Zulen and B Tissot (2019): "The use of big data analytics and artificial intelligence in central banking", *The Capco Institute Journal of Financial Transformation: Data Analytics*, 50th edition, November.

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Paradigms for modern statistical data frameworks and systems[1]

Michal Piechocki,

Eurofiling

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

BIS IFC Satellite Seminar Post-crisis data landscape: micro data for the macro world

16 August 2019 | Kuala Lumpur | Malaysia

Paradigms for modern statistical data frameworks and systems

Michal Piechocki
Board Member | Eurofiling Foundation p.f.
Chairman | Frankfurt Group Technical Workshop

Key study assumptions

- Interdisciplinary design thinking
  - People, data, methodologies, standards, technologies, interfaces
- 6 qualitative case studies of international projects
  - ESMA TR, ECB SDD, EC FDS, NBS CRD, SCM NBFI, SG Banks
- Focus on 3 key contributing factors
- Practical framework of paradigms

Three lenses

Multi-stakeholder collaboration

- How stakeholders are organised?
- How they cooperate?
- How they remain engaged?

Comprehensive data methodologies

- How to create cross-purpose definitions?
- How to reconcile legal and business differences?
- Use of logical description methodologies?

Platforms and standards

- Technical interfaces and platforms?
- Use of international standards?
- Common identifiers?

Micro data: gains and pains, the trade-offs

- Data as originally stored by FIs
- By product, counterparty, transaction…
- Cross-purpose analysis
- Greater insight
- Automated flagging power
- Greater data sets
- Raw data
- Individual data

- Requires FIs input level definition
- Variety of definitions of each
- Purpose-driven regulation
- Greater responsibility
- False positives, false negatives
- Scarce resources
- Does not capture intentions
- Personal data protection

ESMA Trade Repositories Reporting

Multi-stakeholder collaboration

- Reporting is performed by Central Counterparties and Trade Repositories
- Details of derivative transactions are reported to the Trade Repositories which are obliged to provide access to this data to the Competent Authorities

Comprehensive data methodologies

- Tabular description of reporting obligations
- Each field to be reported described by proper data field and/or description on the ITS/RTS level
- Introduction of new legal frameworks based on EMIR allows for reconciliation of definitions on legal acts level

Platforms and standards

- The reporting to TR may use different formats, such as ISO 20022, XML, CSV or FpML messages
- Single standard defined for direct transaction data feed to Competent Authority, as ISO 20022
- In case of EMIR reporting LEI is encouraged, with other identifiers accepted for specific cases only. For instrument reporting ISIN is endorsed

ECB Statistical Data Dictionary

Multi-stakeholder collaboration

- SDD activities performed by European Central Bank
- Introduction of complementary BIRD imitative allows for boarder cooperation within Expert Group, consisting of 10 National Central Banks and 30 Commercial Banks, coordinated by ECB
- The activities are further decided by BIRD Steering Group
- BIRD database is published and periodically updated on dedicated website

Comprehensive data methodologies

- Creating precise definitions of the data through mapping multiple source dictionaries towards single unified dataset
- Each area is managed by one of dedicated ECB subgroups
- Current SDD consist of more than 14 000 elements used across multiple frameworks
- An example of granular framework incorporated into SDD is AnaCredit with nearly 130 unique fields (and 95 data attributes) used for description of detailed information on individual bank loans in the euro area

Platforms and standards

- Single Multidimensional Metadata Model (SMCube) to combine information usually modelled with multiple standards
- Describing multiple identifiers, with LEI and ISIN used as a common denominator for most frameworks

EC Financial Data Standardisation

Multi-stakeholder collaboration

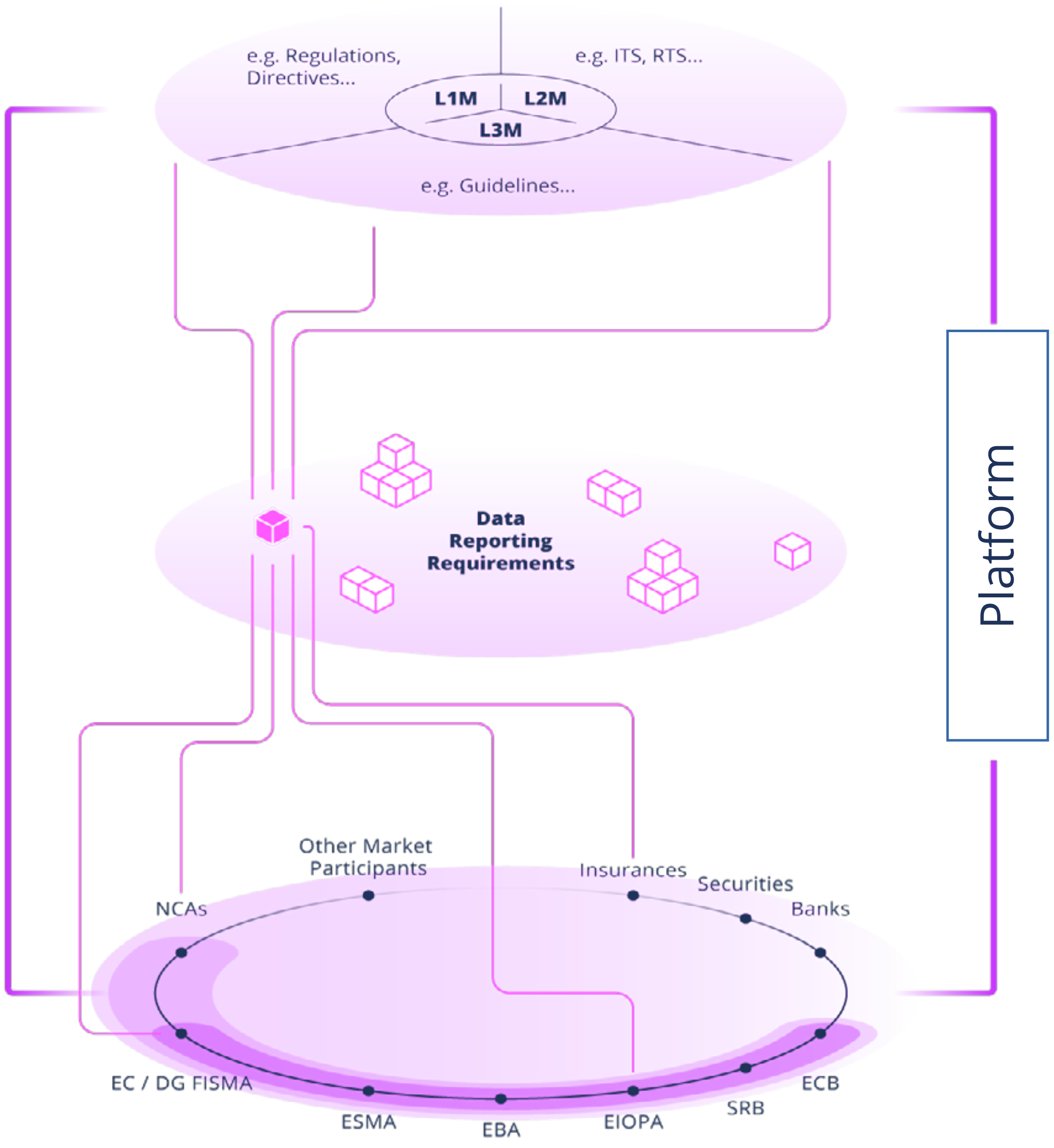
- Supervisory Roundtables (ESMA, EBA, EIOPA, ECB, SRB)
- Lead by European Commission DG FISMA
- Subject-matter experts for data modelling
- Small group of experts for review

Comprehensive data methodologies

- Common definition of DRR
- One data description methodology for information requirements: Data Point Model
- 23 structured supervisory frameworks described
- Direct link to L1M, L2M, L3M legal acts and definitions

Platforms and standards

- 9000+ DRRs
- 33 identifiers
- 7 reporting standards
- All models available on internal collaborative, metadata description platform used by subject matter experts


e.g. Regulations,
Directives...
e.g. ITS, RTS...
L1M
L2M
L3M
e.g. Guidelines...
Data
Reporting
Requirements
Platform
Other Market
Participants
NCAs
Insurances
Securities
Banks
EC / DG FISMA
ESMA
EBA
EIOPA
SRB
ECB

NBS Credit Registry Dictionary

Multi-stakeholder collaboration

- National Bank of Spain leads the effort
- The purpose: enable single standard for data collection
- Development started around credit registry dictionary
- Public consultation of taxonomies

Comprehensive data methodologies

- Definitions started from loan-by-loan system and extended EBA (COREP/FINREP), BSI-MIR
- Single dictionary
- This allowed to extend into FINREP solo
- Methodology: Data Point Model
- Close tie to instructions / circulars (legal)
- Key data owners involved in definition of dictionary
- Top institutional leaders led the proiject

Platforms and standards

- Standards: SDMX, XBRL
- BdE developed an open source software

SCM Private Retirement Schemes

Multi-stakeholder collaboration

- Fund management industries
- The purpose: standardised data collection
- Templates consulted with stakeholders
- Simple interface for data collection

Comprehensive data methodologies

- Methodology: Data Point Model
- Supervisor develops a single understanding of data requirements
- Open tables for threshold-restricted top transactions

Platforms and standards

- Standard of exchange: XBRL
- Best practices driven by experiences of the Bank of England
- Considerations of reuse of other regulatory dictionary for financial reporting

Singaporean Banking Initiative

Multi-stakeholder collaboration

- The purpose: to develop common understanding of supervisory and statistical MAS data requirements
- Lead by a group of private vendors and consultancies
- Founded by 8 FIs in 2016, 32 FI as of 2019
- FIs and Big4 comment and provide feedback on common draft models
- Regulator allowed to partially recover costs through grants

Comprehensive data methodologies

- Methodology: Data Point Model
- Single data dictionary covering MAS 610 and 1003 requirements
- Reconciliation of differences through expert review
- Updated in iterations based on changes in legal requirements and system constraints
- Fis can extend the common dictionary to match their internal requirements

Platforms and standards

- Key output: XBRL taxonomy
- Interoperable with SQL, JSON, XML, CSV
- Subject matter experts use common metadata modelling platform for design and review of data models

Observations

- Five initiatives define requirements for micro data
- Five initiatives used a common open data design methodology: Data Point Model
- Four initiatives emphasized early engagement of relevant stakeholders in design phases of data requirements
- Two initiatives relied almost entirely on data requirements as defined by industry participants
- Two initiatives implemented collaborative platforms for metadata design

Observations (continued)

- One initiative relied on a shared infrastructure for data collection
- Three initiatives relied on common international identifiers: LEI, ISIN (ISO standards)
- Five initiatives relied on open XBRL data collection standard
- Six initiatives facilitate linked business rules

Framework of paradigms

- Collaboration with stakeholders
- Common data understanding
- Comprehensive data models
- Linked business rules
- Agility of data methodology
- Platforms supporting continuous feedback
- Technical interoperability
- Balance between policy and market practice

Thank you

Contact

Michal Piechocki
michal@eurofiling.info

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Data management in the data evolution era at Bank of Thailand[1]

Pimpimol Chansang,

Bank of Thailand

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Data management in data evolution era at Bank of Thailand

Pimpimol Chansang[1]
Bank of Thailand

Abstract

At present, Thai economy rapidly changes and faces with Volatility, Uncertainty, Complexity and Ambiguity (VUCA) environment. At the same time, these cause the proliferation of data sources, in particular, the surge in demand for new data and profound analytics. To implement a policy more efficient, traditional data may not adequate for Bank of Thailand (BOT) to make a policy decision. Therefore, the BOT has adjusted and paved the way for reaching to the economic challenges by setting direction and scope in the BOT 3-year strategic plan (2017-2019). For data management area, the BOT increases in capacity to analyse data at micro level and to make a greater use of high frequency data from various sources in order to complete more coverage of economic situations. The main objective of this paper is to portray the current state of data management at the BOT under the data evolution era. The BOT has made consideration efforts to improve the internal data management. Consequently, more granular data is an evolution for timelier and richer policy analysis in this time. One of the BOT's studies uses corporate loans to study the information-based lending instrument to make its standards viable. Nowadays, various important data are obtained from both government and private organizations. The BOT has collaborated with various organizations because the data and the academic cooperation are provided appropriately to meet the needs and benefits for the mutual organization. However, micro data has an enormous potential to help policymakers overcome several constraints while also create challenges at the same time. In particular, the data also contains confidential information on the individual, therefore, it should balance between usage and protection of the data as well as comply with data protection law of the country including international standards.

Keywords: data management, data collection, micro data

[1] Data Management Department, Bank of Thailand. Contact PimpimoC@bot.or.th. The author is grateful to Dr. Somsajee Siksamat, Senior Director at Data Management Department, Bank of Thailand for guidance and valuable comments. The views expressed are my own and do not necessarily represent the Bank of Thailand.

Contents

1. Introduction

In the VUCA world of high Volatility, Uncertainty, Complexity and Ambiguity, economic environment has changed dramatically. Corresponding to the world of SMAC which stands for Social, Mobile, Analytics and Cloud, it also completely changes pattern of individual behaviours. Moreover, the transformation could bring about a shock transmission spread more easily throughout the world. Therefore, the evolution has developed globally and the organizations both private and government sectors have to shift an economic evaluation into various dimensions which gives the new growth of opportunities and challenges.

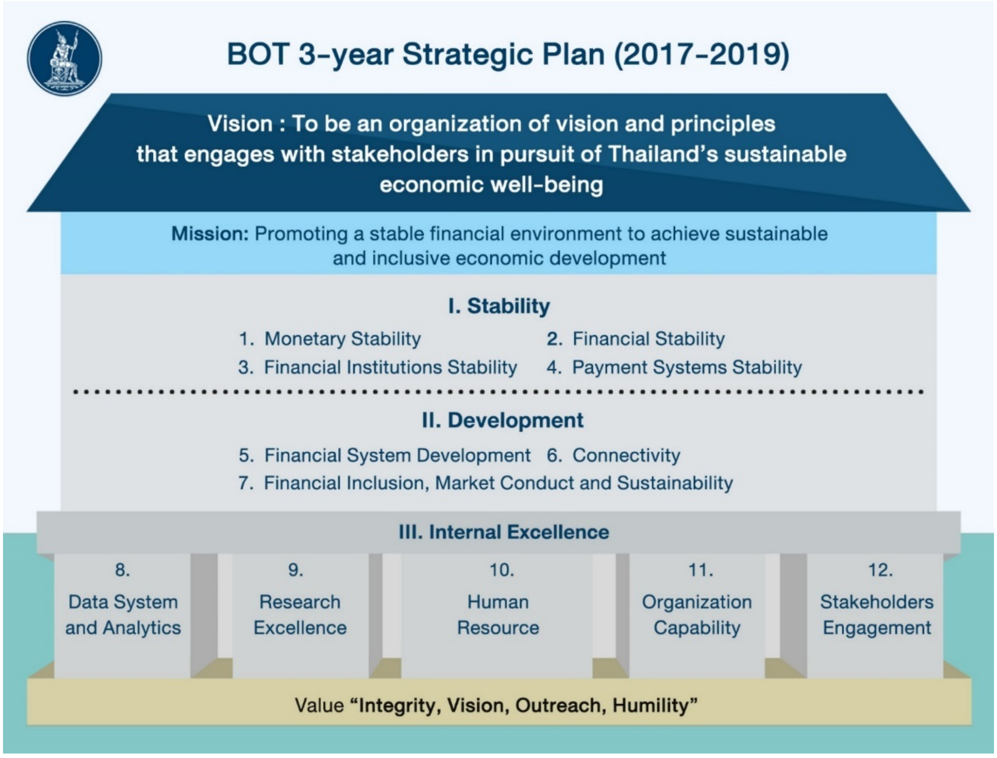
There has been a notable rise in the volume and variety of data which brings the difficult to grasp such a new economic activity by conventional measures. So, an analysis is developed to ensure insightfulness and deeply understanding of the evolution from conventional to unconventional analysis. Data is changed from structured to unstructured data. Methodology and tools level up from macroeconomic models to machine learning to support data analytics in the transformative world. The integration of demand and supply of the data could make an analysis more efficient in terms of deeply and widely analysis which develops from macro to micro view in order to understand individual behaviours and support evidence-based policy. However, these lead to a question that how the data management will develop in the age of disruptions. Then, this study proposes to show the current state of data management at Bank of Thailand consisting of strategic direction of organization developing to support structural reform of the economy. In addition, the paper provides the linkage between supply and demand of data which shifts from analyzing the conventional to unconventional way. Finally, the study portrays the challenges of data coming with the age of data evolution.

2. Data Management in Data Evolution Era

2.1 BOT 3-year Strategic Plan

The BOT 3-year strategic plan comprises of 12 strategies grouped into 3 dimensions (figure 1). The main pillars contain stability, development and internal excellence. The stability of the economy and the financial system is the BOT's core mandate. The development of financial sector is provided to help to promote sustainable and inclusive growth towards digital economy. Finally, internal excellence is a critical foundation and enables both stability and development to functioning efficient.

Figure 1: BOT 3-Year Strategic Plan (2017-2019)



Source: Bank of Thailand

Along with the BOT's strategic plan, the Data Management Department (DMD) also has a plan to support the organization. Firstly, the BOT will enhance capacity on data analytics, especially through the use of micro-level data to complement aggregated data and understanding structural adjustments and managing the organization. Moreover, the BOT will implement the data governance organization-wide including Data Leak Prevention (DLP). The data governance process helps to enhance proper data collection, analysis of data and enhancing further innovative use of data. Secondly, the DMD will acquire new data and improve existing data to meet user requirements and ensure the high-quality of data including timeliness and adequacy of data from various sources. Thirdly, the BOT will widely-make use of data, especially financial and economic data for creating and developing data and statistic indicators, including academic research. Lastly, the BOT will encourage academic cooperation of various organizations in order to set public data bureau in financial and economic data and contribute data sharing as well as create a data standard of country.

2.2 Data Collection Framework

Data in Bank of Thailand is collected in several kinds both traditional and non-traditional data. However, the main sources are financial and non-financial institutions as well as could be separated into four areas consisting of financial institution data, financial market data, payment data and economic data. All of the data are provided to serve financial and economic stability, statistic purposes and monetary policy formulation including financial markets, financial institutions and supervision. For data collection legislation, we abide by the laws and regulations consisting of Bank of Thailand Act B.E. 2485 (1942), Financial Institution Business Act B.E. 2551 (2008), Exchange Control Act B.E. 2485 (1942) and Payment System Act B.E. 2560 (2017). There are various types of data collecting in central bank such as regulated data and data from collaboration, Memorandum of Understanding (MOU), internal data and public data. In addition, there are other data which are also collected from public bodies and private companies in order to complete coverage of data and meet users' needs such as administrative data and survey data.

In the digital world, new data are created every day, therefore, the collection is transformed to next generation of data including big data to monitor economic and financial stability, statistics proposes and policy analysis. There are various types of unconventional data such as micro data and big data acquiring in respond to rapidly change in society and gain wider and deeper insight into individual behaviors and social patterns. They also complement and expand data that are collected conventionally. However, the data in the organization is usually collected separately. Therefore, to operate and manage the data efficiently and complete the high quality of data, the technology of integration platform should be provided to support the process of validating, cleansing and linking data from different sources in order to enhance data usage in many aspects. However, we try to create more value of data by linking between data points from various data sets through the platform, namely individual data integration system (IDI). For example, there is linking between loan data and export-import data in order to complete an analysis in the new structure of economic situation. This integration could help to encourage analyzing data from various sources with many aspects and more profoundly. Regarding the data is in form of the micro data and transactional base, then, data anonymisation and data masking are implemented ensuring the data privacy and security protection.

2.3 Micro Data Usage

In response to rapid changes, the BOT have to upgrade analysis by using the micro data which can both directly and indirectly benefit to macroeconomic analysis. This is a way of change from conventional to unconventional analysis in order to fill the data gap and increase the usage of micro data together with the use of new data analytics techniques. These, in turn, help to support the needs of statistical data for analysis aligned with the data strategic plans as well as elevate the quality and standard of the statistical data in a new economic landscape.

Basically, an analyst usually starts with a financial balance sheet in order to monitor asset and liability of financial sector. However, this is not enough for policymaker to make the policy decision. This is the reason why the organization has to use micro data for macro analysis. An information-based lending measure is implemented to better access of financial resources. Pinthong et al. (2018) studied corporate debt-level data and found that a revenue-based lending has limited access to bank loans of SMEs, especially new firms, low value fixed assets firms and low income revenue firms. Therefore, we allow financial institutions to use of alternative data in addition to income such as mobile data usage, utility usage and bill payment in determine the ability to pay of the borrowers which provides the greater benefit of financial access to SMEs. Moreover, according to credit bureau data, the results from the study of Chantarat et al. (2017) also showed that Thais load up on debt at a younger age and owe it longer. This led to implement the regulations on credit card and personal loan to manage household debt of the country.

For property market, the BOT has conventional statistics data to monitor real estate conditions such as construction areas permitted data, Residential Property Price Index (RPPI) and Land Price Index (LPI). However, to capture macro analysis by using only aggregated data may be insufficient, therefore, Dummee and Trongthamakit (2018) incorporated micro data such as electricity usage data to expand the coverage of analysis in real estate market. This analysis can help to monitor bubble in the property market by compiling an occupancy rate from the number of unit in condominium representing real living behaviours of Thai people. In addition, the data also provides benefit for evaluating manufacturing and tourism sector development from electricity usage. Similarly, Samritpiam and Wanitthanankun (2018) applied bulk payment data to construct Employment Revenue Index (ERI) to monitor labor market both trend of number and income of labor for employment conditions.

The BOT also seeks to take advantage of big data which potentially enhances making policy decision. Sawaengsuksant (2018) used Google Correlation to compliment monitoring economic conditions such as consumption, purchasing power, consumer confidence and number of unemployed persons. These information can enhance predictive power of forecasting and supplement the traditional data. Moreover, the BOT also stepped up in analysis by using text analytics in order to support policy formulation and internal excellence. For example, the BOT used text analytics to improve operational efficiency of statistics compilation such as classification of government expenditures and analyze employees' health as well as explore minutes of commercial banks to help understand behavior and culture of the banks. Therefore, this is a way of change how the micro data can serve the policy analysis in the macro world and also make policy better and more efficient in this data evolution era.

3. Challenges

Along with the process of acquiring, processing, keeping and using data, there are several challenges must be overcome before we transform into a digital organization. There are three main tasks consisting people, technology and laws and regulations. Firstly, people are a key driver to deliver work and make data speak. Thus, it is necessary to develop analytical skills in order to tackle with the new things both data and tools. For example, data scientists, lawyers, IT specialists should work alongside statisticians and economists in order to fill the gap to each other. To level up capacity of managing and using unconventional data, therefore, collaborating with abundant human skills is an important to work across their boundaries and could take advantage from this age of data. Secondly, technology data platform is preferred to support data management from collecting data until using data both conventional and unconventional data in easy way. The data platform should provide technical options to complete the quality of data in terms of cleansing, validating and publishing. Moreover, the technology platform should support data integration which can combine various data sets particularly granular data and big data-sets. At the same time, the technology should apply Data Leak Prevention (DLP) to safeguard privacy data and ensure that data is not exposed to anyone who has no right to access it. Finally, laws and regulations directly affect to the data management in the organization such as the Personal Data Protection Act B.E. 2562 (2019), Thailand Cybersecurity Act B.E.2562 (2019), the Digital Government Act B.E. 2562 (2019) and the General Data Protection Regulation EU 2016 (GDPR). Since the micro level data usually contains large amount of sensitive and personal information, any violation of the privacy and confidentiality risks are a major concern. Moreover, data usage is not limited only in the organization or in the country but it can share across country to enhance data usage. If there is insufficiently protected information, the loss of data can lead to reputational damages and loss of trust in the organization. Data governance policy, therefore, is an important role in governing data management and encouraging aware of accountability as well as responsibility to the data usage in the organization.

4. Way Forward

Under uncertainty situations, data becomes a significant tool for analysing and making better policy decision. To make the BOT's work smooth and insist on this environment, then, the way forward and preparedness plans are followed. Business use case should be implemented first to acquire new data. This leads us to know what kind of data can serve us and what policy we want to do. If we have the right direction, we will be able to solve problem clearly, meet ultimate goal with timely and take it into practice promptly. It also helps to balance the use between demand and supply of data. Second, data platform should be deployed to operate and manage the data efficiently in terms of capacity, quality and use. Furthermore, Data Leak Prevention (DLP) technology will be used to protect confidential and privacy data in terms of keeping and sharing data as well as to align with the data governance policy. To provide more improving the analytical capacities of the organization, we are studying such a suitable technology to integrate various large micro data from different data sources. Moreover, the technology should support appropriate visibility both masked and unmasked confidential data depending on roles and responsibilities as well as purposes. Third, the BOT implemented Data Governance Policy (DGP) on July 2019, which provides a guideline for proper data management and helps to increase and sustain data governance awareness and accountability at all levels in the organization. Consequently, the BOT is conducting a code of conduct which provides a practical guideline for managing data and to assure that all entities could govern data under the same standard. In the same way, central data registry is prepared in order to centralize and contain all data entities with systematically. It also provides data classification which helps to scrutinize the confidentiality of data and support role-based access control as well as to facilitate effectiveness of data management. Last, data should be fully utilized in the organization including across country. At the same time, the data will be kept securely and confidentially along with compliance with the laws and regulations of the country and international standards. If data is distributed, this could help not only reduce the burden on data providers but also contribute abundant knowledge to each organization and support data ecosystem of the country.

References

Bank for International Settlements (2015), "Central banks' use of and interest in big data". https://www.bis.org/ifc/publ/ifc-report-bigdata.pdf

Bank for International Settlements (2016), "The Sharing of Micro Data – A Central Bank Perspective". https://www.bis.org/ifc/publ/ifc-report-microdata.pdf

Bank of Thailand (2016), "Bank of Thailand's 3-Year Strategic Plan 2017-2019". https://www.bot.or.th/English/AboutBOT/RolesAndHistory/Documents/3yearsStrategicPlan.pdf

Bank of Thailand (2019), "Annual Report 2018". https://www.bot.or.th/English/ResearchAndPublications/Report/DocLib_AnnualEconReport/AnnualReport2018_En.pdf

Bellini, G. Cecconi, N. Gaetano, L.D., Monetti, F., Papa, P., and Ranaldi, R. (2018), "Centralizing Data Collection Implementation: The Istat Experience". Workshop on Statistical Data Collection. https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.58/2018/mtg7/DC2018_S0_Papa_etal_Italy_AD.pdf

Chantarat, S., Lamsan, A., Samphantharak, K., and Tangsawasdirat, B. (2017), "Thailand's Household Debt through the Lens of Credit Bureau Data: Debt and Delinquency". https://www.pier.or.th/wp-content/uploads/2017/08/pier_dp_061.pdf

Chen, H., Chiang, R.H.L., and Storey, V.C. (2012), "From Big Data to Big Impact". MIS Quarterly. https://pdfs.semanticscholar.org/f5fe/b79e04b2e7b61d17a6df79a44faf358e60cd.pdf

Dummee, J., and Trongthamakit, K. (2018), "Condominium Occupancy Rate in Bangkok and Its Vicinity from Electricity Meter Data Analytics". https://www.bot.or.th/Thai/Statistics/Articles/Doc_Lib_statisticsHorizon/Electricity.pdf

European Central Bank (2017), "Policy analysis with big data". Speech at the conference on economic and financial regulation in the era of big data. https://www.ecb.europa.eu/press/key/date/2017/html/ecb.sp171124.en.html

Hammer, C., Kostroch, D.C., and Quiros, G. (2017), "Big Data: Potential, Challenges and Statistical Implications". https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2017/09/13/Big-Data-Potential-Challenges-and-Statistical-Implications-45106

International Monetary Fund (2016), "Behind the Scenes with Data at the IMF an IEO Evaluation". https://www.imf.org/ieo/files/completedevaluations/Statistics%20-%20Full%20Report.pdf

Musigchai, C. (2008), "Relationship with survey data providers: the Bank of Thailand's experiences". https://www.bis.org/ifc/publ/ifcb28zzj.pdf

Paweenawat, A., J. Pinthong, K. Samphantharak, and N. Wasi. (2018), "Corporate Debt in Thailand: What We Learn from the National Credit Bureau Data". Discussion Paper, Puey Ungphakorn Institute for Economic Research. https://www.pier.or.th/?abridged=มุมมองสินเชื่อธุรกิจไท-2

Samritpiam, K. and Wanitthanankun, J. (2018), "Tracking Formal Employee Salary from Payment Data".
https://www.bot.or.th/Thai/Statistics/Articles/Doc_Lib_statisticsHorizon/Bulk.pdf

Sawaengsuksant, P. (2018), "Standardised approach in developing economic indicators using internet searching applications".
https://www.bis.org/ifc/publ/ifcb50_18.pdf

Tissot, B. (2017), "Big data and central banking".
https://www.bis.org/ifc/publ/ifcb44_overview_rh.pdf

Tissot, B. (2018), "How can big data support financial stability work?".
https://unstats.un.org/Unsd/nationalaccount/workshops/2018/rio/BIS.pdf

Wasi, N., P. Bumrungruan, C. Monchaitrakul, M. Rudtanasudjatum, and K. Samphantharak. (2018), "SME Debt in Thailand: A Perspective from Loan-Level Data". Discussion Paper, Puey Ungphakorn Institute for Economic Research.
https://www.pier.or.th/?abridged=มุมมองสินเชื่อธุรกิจไท

Wibisono, O., Ari, H.D., Widjanarti, A., Zulen, A.A., and Tissot, B. (2019), "The use of big data analytics and artificial intelligence in central banking".
https://www.bis.org/ifc/publ/ifcb50_01.pdf

ธนาคารแห่งประเทศไทย
BANK OF THAILAND

The IFC-ECB-BNM Satellite Seminars

Data Management in the Data Evolution Era at Bank of Thailand

Pimpimol Chansang

Data Strategy Team, Data Management Department, Bank of Thailand

16 August 2019

ธนาคารแห่งประเทศไทย
BANK OF THAILAND

Outline

1. Introduction of Data Evolution Era
2. Objective of the study
3. Data Management
4. Challenges
5. Way Forward

Data Evolution Era

Data is one of the strategic asset and can make many benefits for an organization.

Structured Data

- Excel
- Time Series Data
- Sales
- Surveys
- Financial Statement

Unstructured Data

- Chat
- Customer/Member Transactions
- Notes & Text Fields
- Email
- Social Media
- Voice Transcriptions
- Call Center
- Online Communities
- Ratings & Reviews

Technology

- cloudera
- hadoop
- API Application Programming Interface

Tools

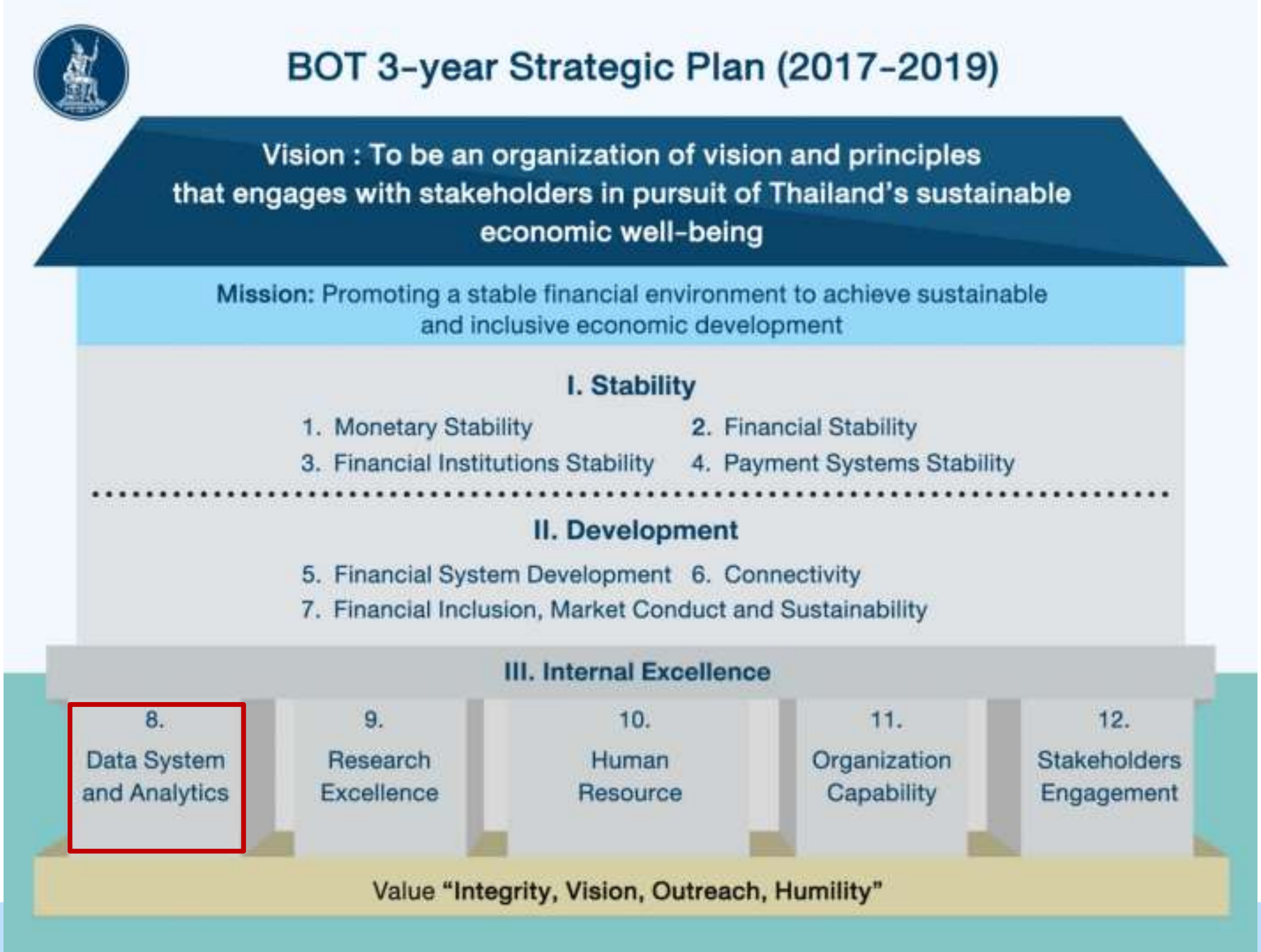
- EViews
- Stata
- R
- MATLAB
- tableau SOFTWARE
- sas
- python

Analysis

- Analysis Using Micro Level Data
- Prediction
- Clustering
- Text Mining
- Network Analysis

Objective of the study

- To portray the current state of data management
  - BOT's 3 Year Strategic Plan
  - The Journey of Micro Data
  - Micro Data Usage
- To state challenges associated with data management

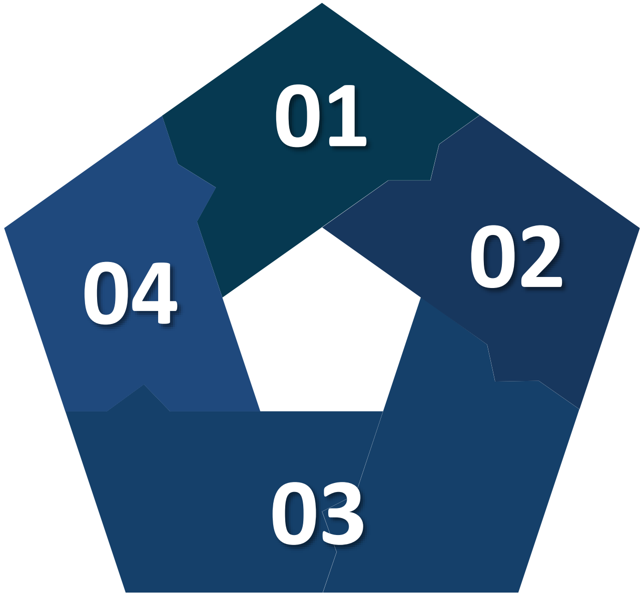

BOT 3-year Strategic Plan (2017-2019)
Vision : To be an organization of vision and principles that engages with stakeholders in pursuit of Thailand's sustainable economic well-being
Mission: Promoting a stable financial environment to achieve sustainable and inclusive economic development
I. Stability
1. Monetary Stability
2. Financial Stability
3. Financial Institutions Stability
4. Payment Systems Stability
II. Development
5. Financial System Development
6. Connectivity
7. Financial Inclusion, Market Conduct and Sustainability
III. Internal Excellence
8. Data System and Analytics
9. Research Excellence
10. Human Resource
11. Organization Capability
12. Stakeholders Engagement
Value "Integrity, Vision, Outreach, Humility"

ธนาคารแห่งประเทศไทย
BANK OF THAILAND

Strategic Direction 2017-2019: Data System and Analytics

1. Enhance capacity on data analytics in assessing economic conditions & to support strategic direction and managing organization

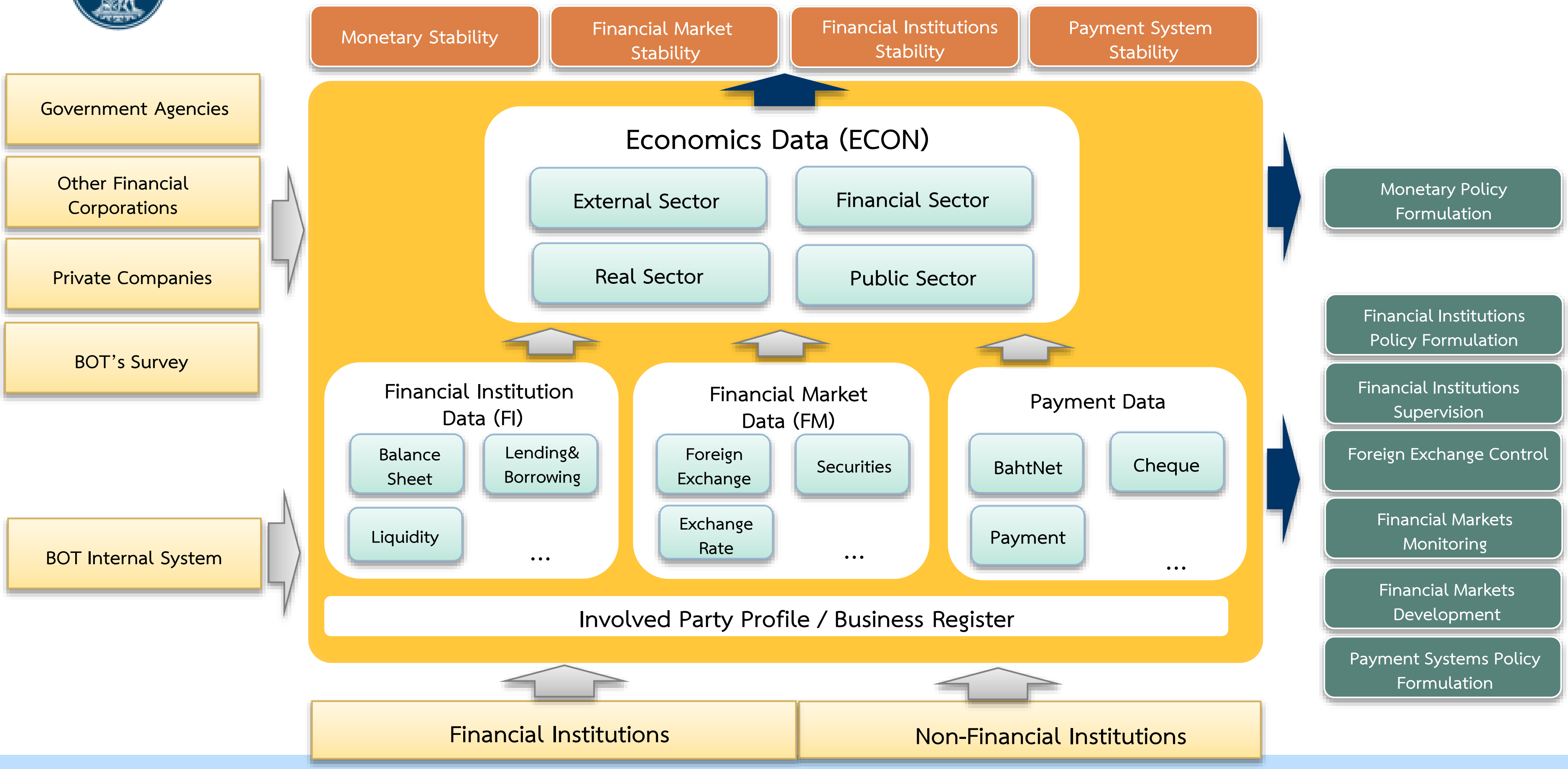
2. Acquire data from various sources especially near real-time data to meet timeliness, adequacy and quality of data



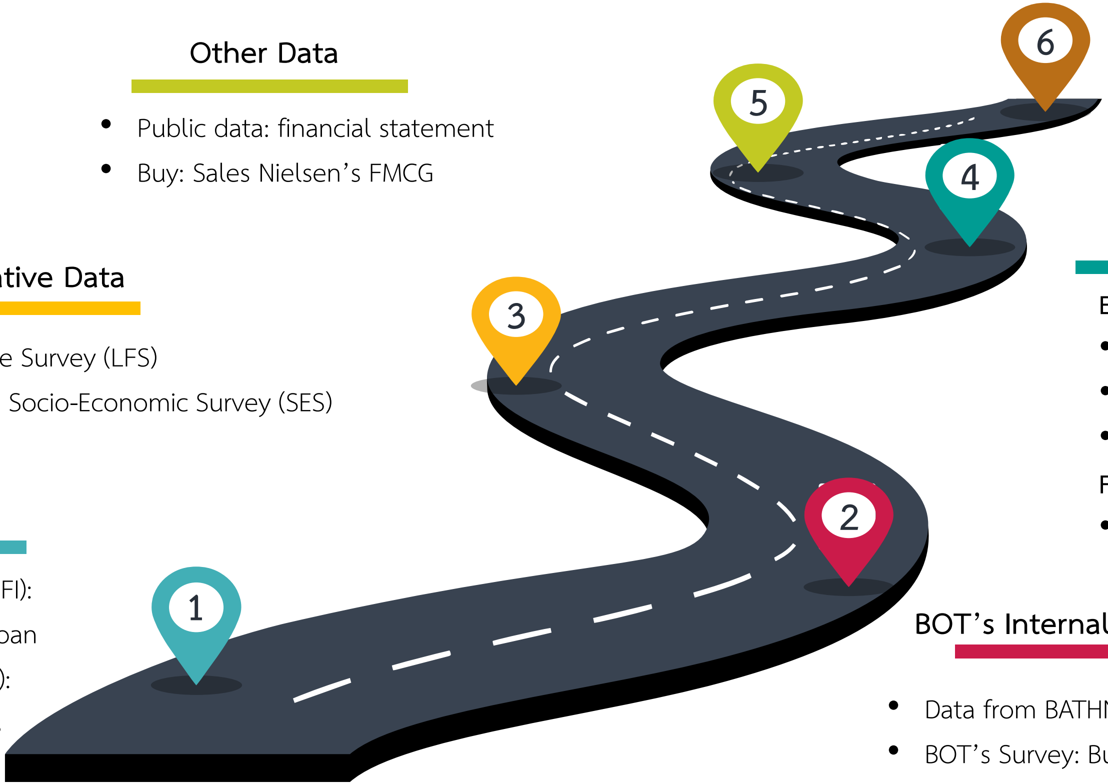
3. Widely-made use of data, especially financial and economic data

4. Contribute to set public data bureau in financial and economic data

Current State of Data



The Journey of Micro Data



Regulated Data

- Financial Institution Data (FI): business loan, mortgage loan
- Financial Market Data (FM): FX transanction, securities
- Payment Data: interbank bulk payment

BOT's Internal Data/Survey

- Data from BATHNET system
- BOT's Survey: Business sentiment survey, IIP

Collaborative Data

- Labor Force Survey (LFS)
- Household Socio-Economic Survey (SES)

MOU

Economic Data

- Electricity consumption
- Employment data
- Import-export data

Financial Data

- Credit Information

Other Data

- Public data: financial statement
- Buy: Sales Nielsen's FMCG

Next Gen Data

- Big Data

Micro Data Usage in Data Evolution

Analytically changed from conventional to unconventional data and use, which can leverage both directly and indirectly benefit to macroeconomic and financial statistics, and finally policymaking.

1 Information Based Lending Policy

- To use corporate debt-level data in analyzing the development and the current situation of corporate debt in the country
- A revenue-based lending has limited access to bank loans of new firms, low value fixed assets & low income revenue firms
- Changing from conventional data to behavioral data: moile data, utilities serives payment for credit approval
- Allows SMEs to better access financial resources

2 Property Market Analytics

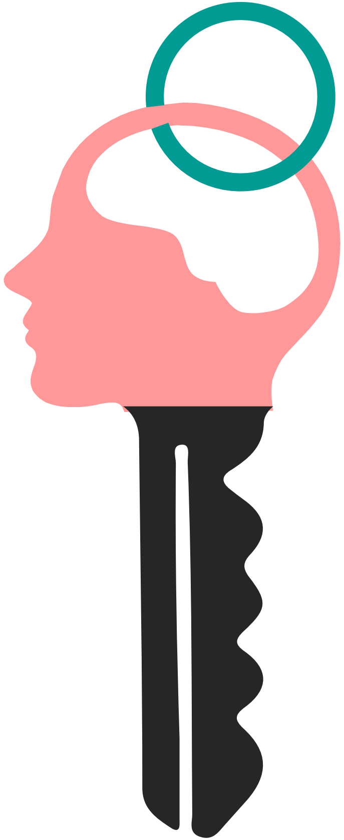
- **Conventional Indicators**: Construction Area Permitted Indicator, Residential Price Property Index, Land Price Index
- **Using across with unconventional data**: electricity consumption from electricity meter on monthly basis
- To compile condominium occupancy rate in Bangkok and Its vicinity
- In constructing occupancy rate to analyze consumer behavior of real living in condominium and monitor properties market
- Sopporting to analyze real estate conditions and tourism from electricity usage data

3 Use text analytics to improve operational efficiency

- To classify government expense by code number to support a compilation of public finance statistics
- To analyze health profiles/records of BOT's employees
- To explore minutes of commercial banks using text analytics to help understand behavior and culture of the banks
- To analyze BOT's notifications using text analytics to support RIA (Regulatory Impact Analysis)

Reference: 1. Paweenawat, A., J. Pinthong, K. Samphantharak, and N. Wasi. (2018): "Corporate Debt in Thailand: What We Learn from the National Credit Bureau Data" 2. Knuntip and Jittima (2018) "Condominium Occupancy Rate in Bangkok and Its Vicinity from Electricity Meter Data Analytics" 3. Bank of Thailand Annual Report (2018)

Challenges



People

- New Skill Profiles: new (big) data, analytics tools
- Mixed skill: economist, statistician, data scientist, lawyer and business skill
- Capacity: data management and data usage

Technology

- Acquisition: structured & unstructured data
- Process: cleasing, quality, validation
- Keep: linking data to enhance usage

Laws and Regulations

- New Data Laws: do not break the laws

- The Personal Data Protection Act (B.E.2562 (2019))
- The Cybersecurity Act (B.E.2562 (2019))
- The General Data Protection Regulation (GDPR) 2018

- Data Governance Policy

ธนาคารแห่งประเทศไทย
BANK OF THAILAND

Way Forward

1 Business Used Cases

- Collaboration demand and supply of data
- To acquire the right data

2 Data Platform

- Operating and managing: big data, quality, use
- Processing data: masked data vs. unmasked data
- Private cloud storage
- Data Leak Prevention

3 Data Governance

- BOT's Data Governance Policy
- Improving DG Awareness: accountability
- To have code of conduct applied all units

4 Data Collaboration

- Data coordinating in and out organization & cross country
- Help reduce burden of data providers

ธนาคารแห่งประเทศไทย
BANK OF THAILAND

Q&A

Thank You

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

OeNB's reporting data model as RegTech/SupTech solution[1]

Johannes Turner,

Central Bank of the Republic of Austria

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

OESTERREICHISCHE NATIONALBANK
EUROSYSTEM

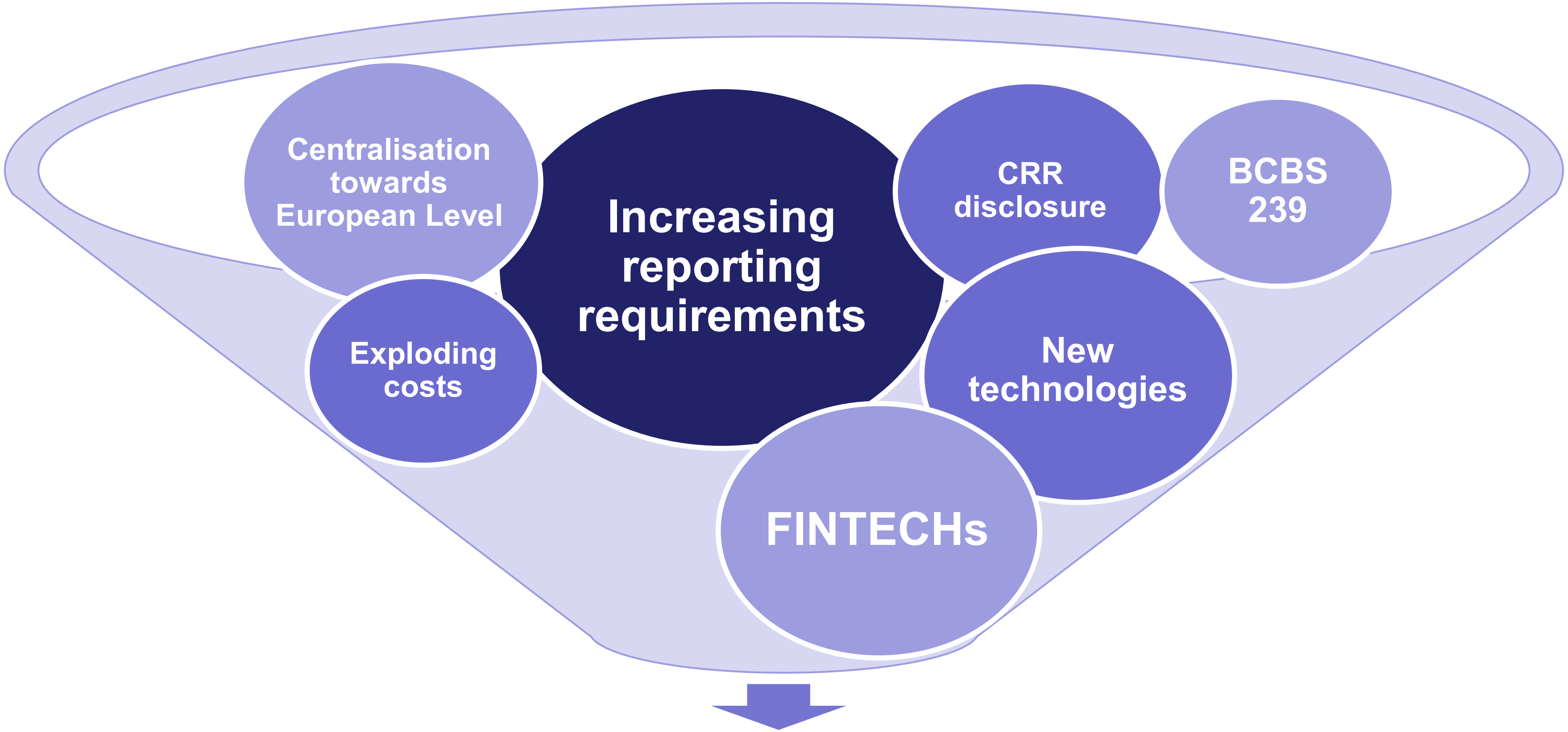
OeNB’s reporting data model as RegTech/SupTech solution

IFC-BNM-ECB Satellite on "Post-crisis data landscape: micro data for the macro world"

August, 16, Sasana Kijang, Bank Negara Malaysia

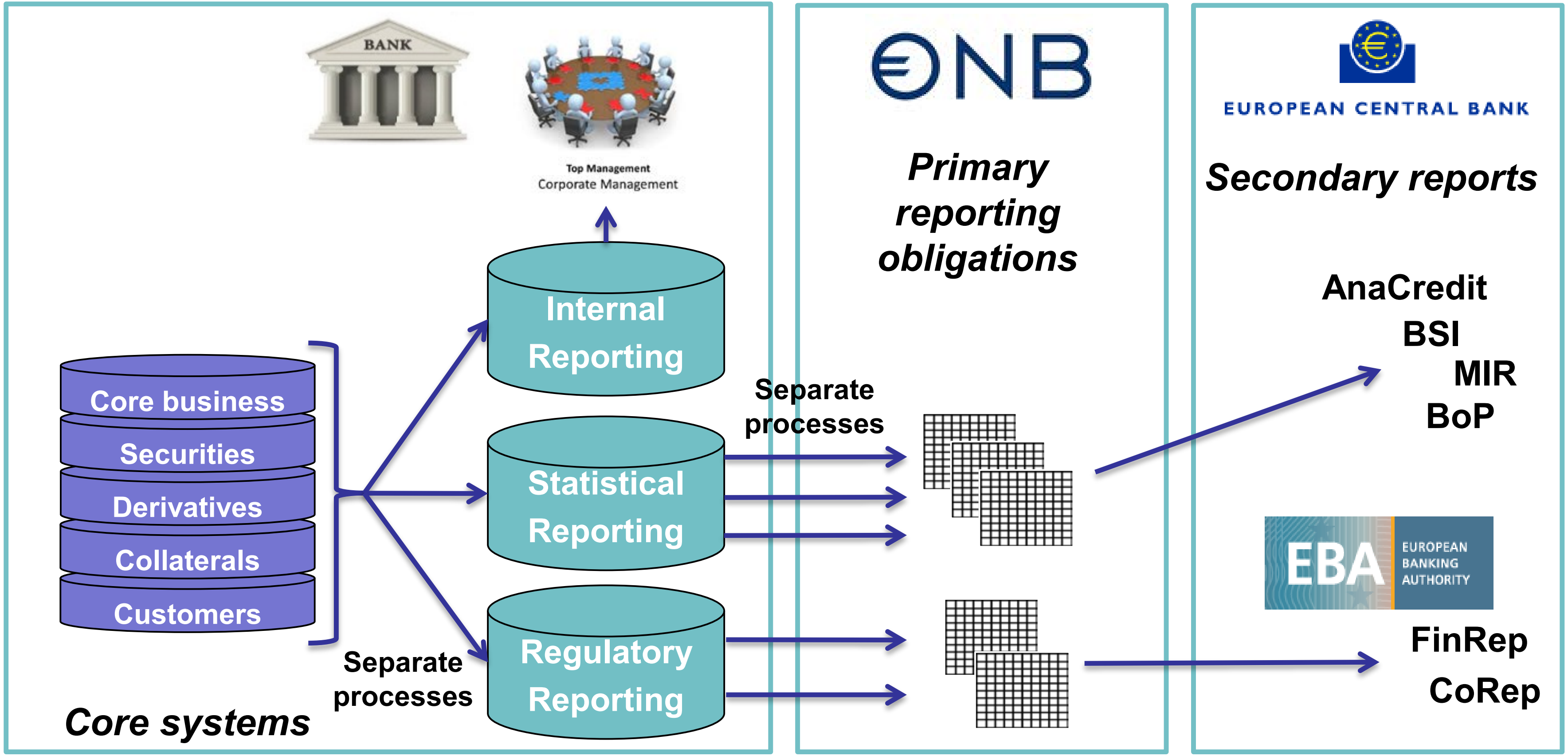
Johannes Turner, Director Statistics Department

Challenges in banks' reporting



Are (central) banks well prepared?

Traditional processes of banks' reports …



… often lead to …

Redundancies

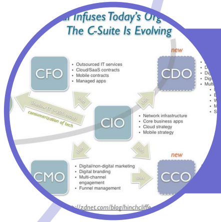
Inconsistencies

Inflexibilities

The RegTech/SupTech perspective

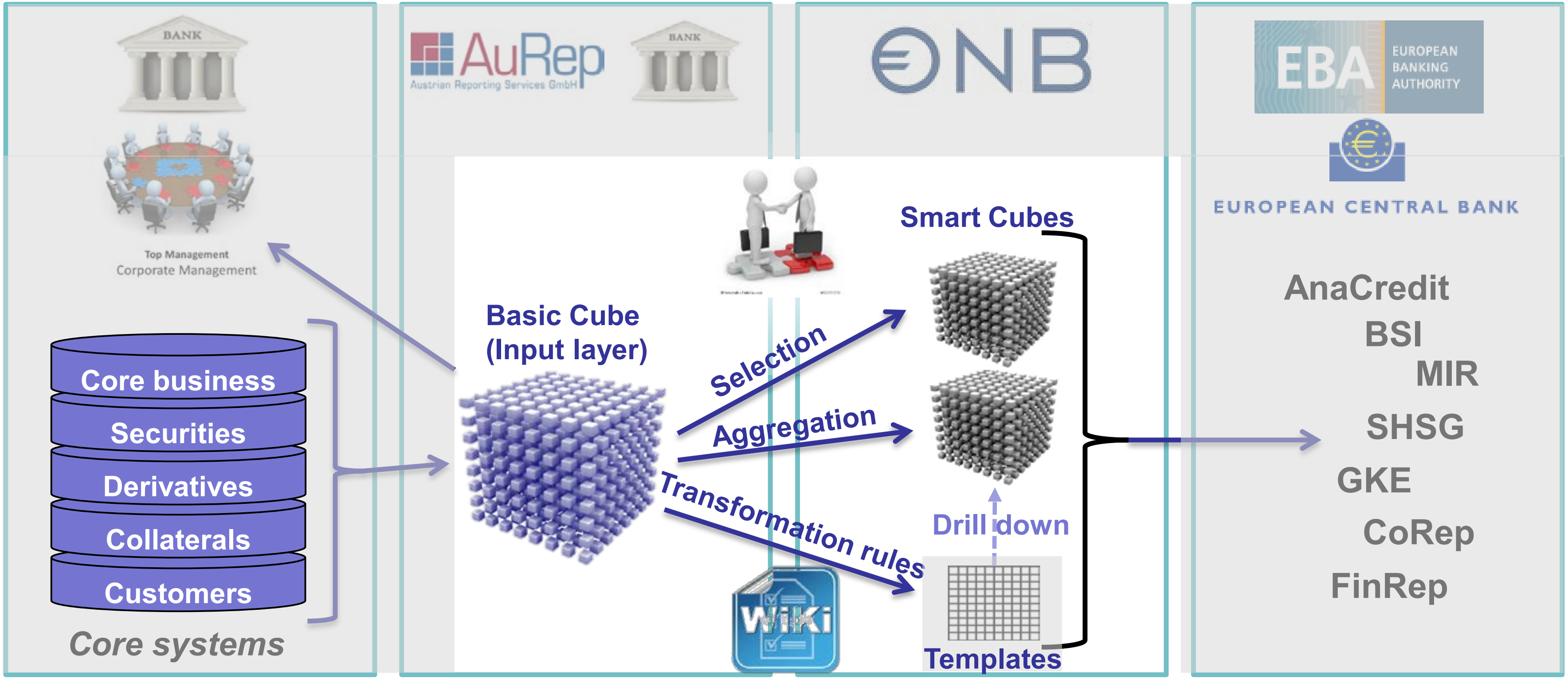
"A combination of technologies and ***innovative processes*** are deployed to modernize data gathering and data analytics"

"It helps supervisory agencies ***digitise reporting*** and regulatory processes. Suptech could be a ***game-changer*** in efficient reporting"



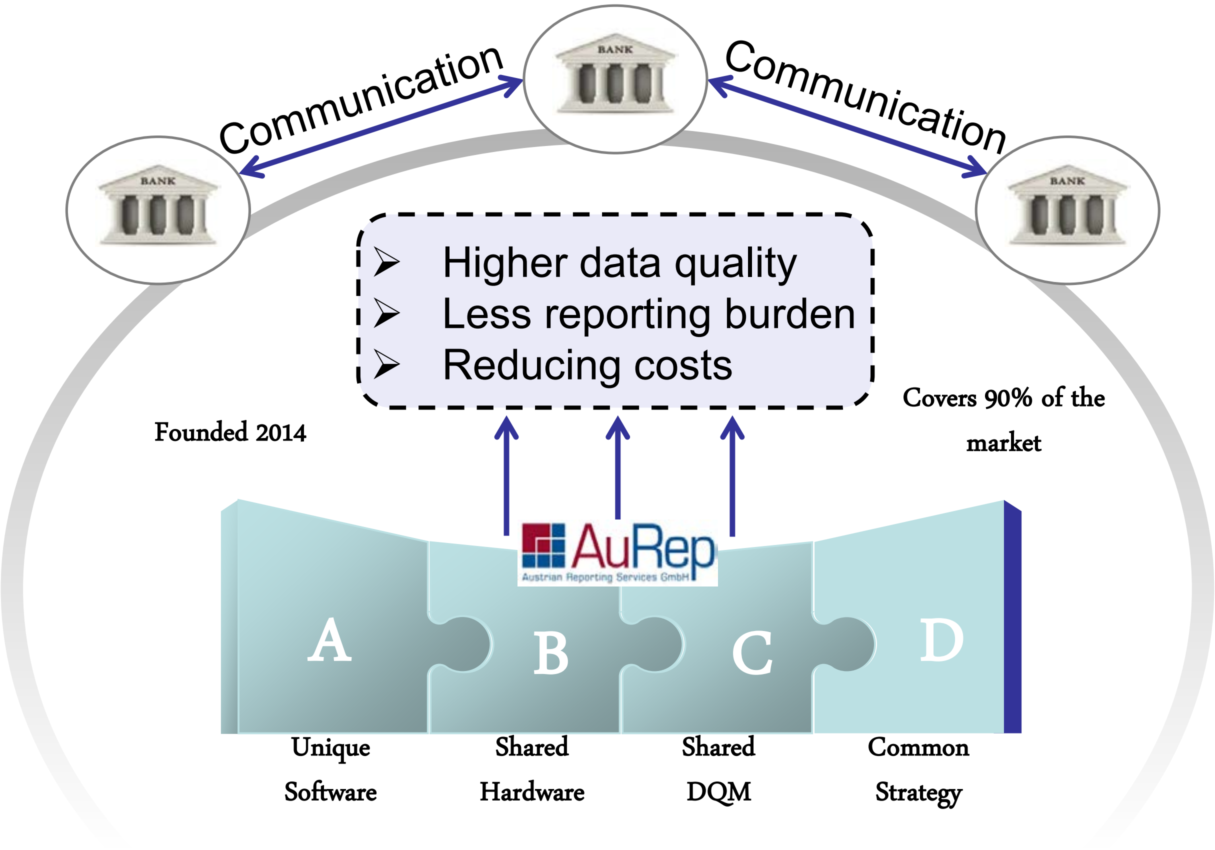
"RegTech/Suptech could have significant ***organisational impact***"

Make reporting more efficient …



… by using an integrated reporting process

Realise synergizes and reduce costs



Use granular data for aggregation and drills

Security-by-security information based on a long term experience since 1991

Sec-by-sec system

ISIN/Non-ISIN

Issuer Country Maturity ... more than 60 attributes

Products

- Balance sheet items statistics
- Securities issues statistics
- Securities holdings statistics
- Insurance corporations statistics
- Pension funds statistics
- Investmentfunds statistics
- Balance of Payments
- Financial accounts
- International Investment Position

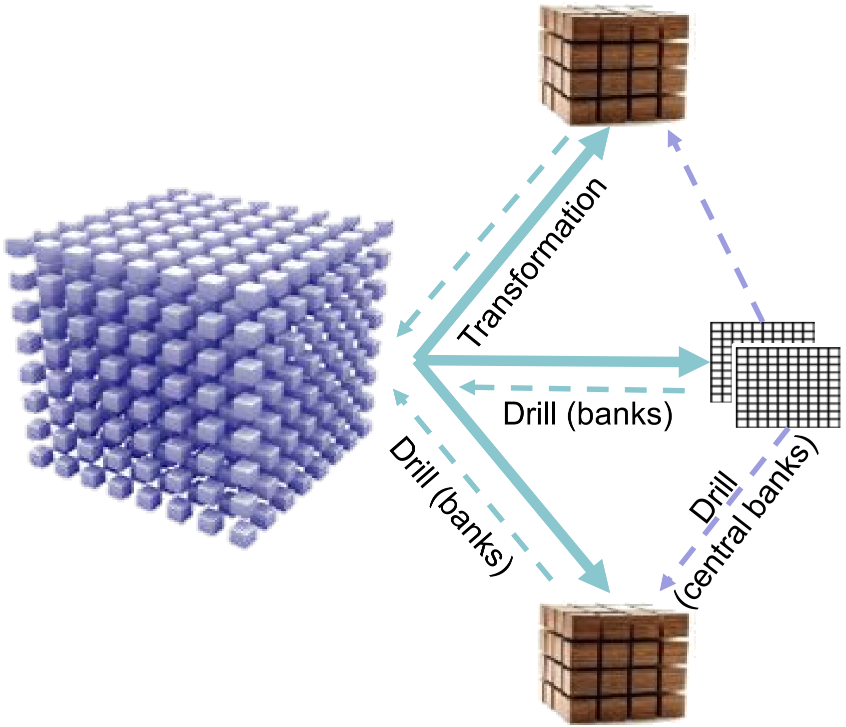
Analysis

- Ad-hoc requests
- Reports
- Press information
- Plausi checks

Production

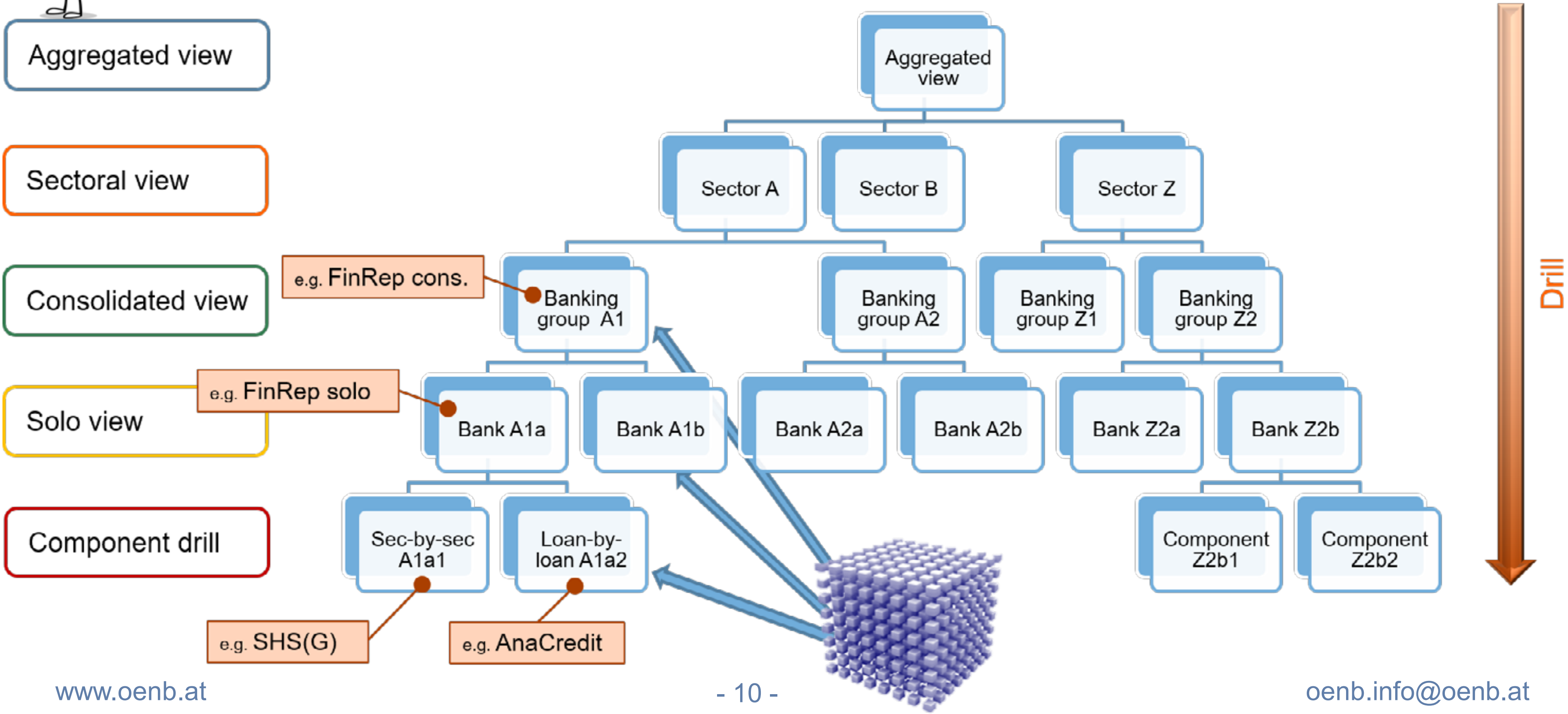
Drill

Apply advanced drill downs for DQM and analytics

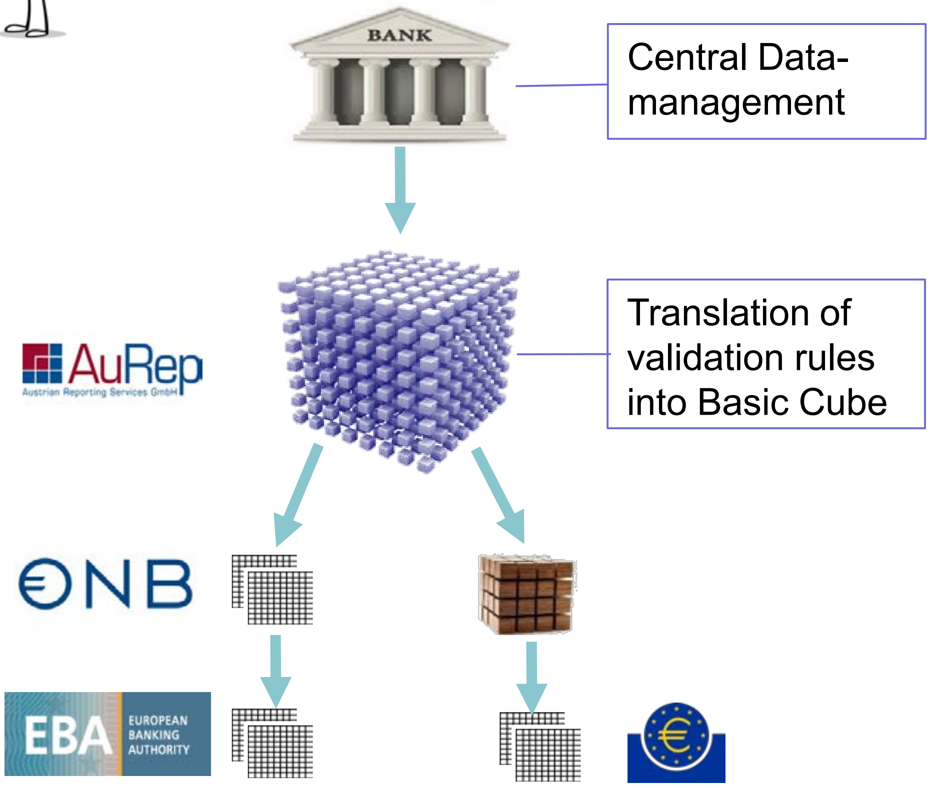


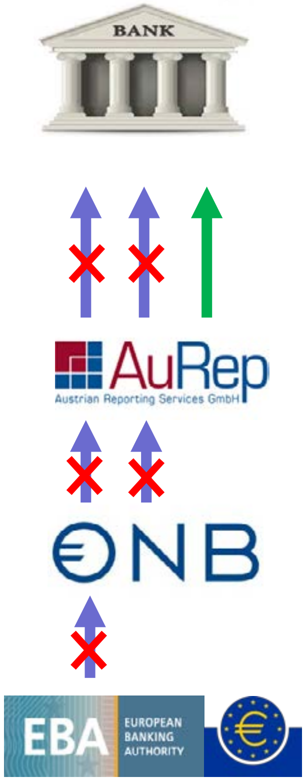
- ➢ Reverse engineering of transformation rules from basic cube to reports allows
  - drills within banks (e.g. from reports to basic cube) and
  - in the central bank (between reports in different granularity, e.g. FinRep to AnaCredit/ SHSG), and
  - timely replies to ad hoc requests
- ➢ Drills show cross linking,
- ➢ Foster understanding, and
- ➢ Minimise enquiries from authorities

Linking different levels of aggregation



Do data quality management at the basis





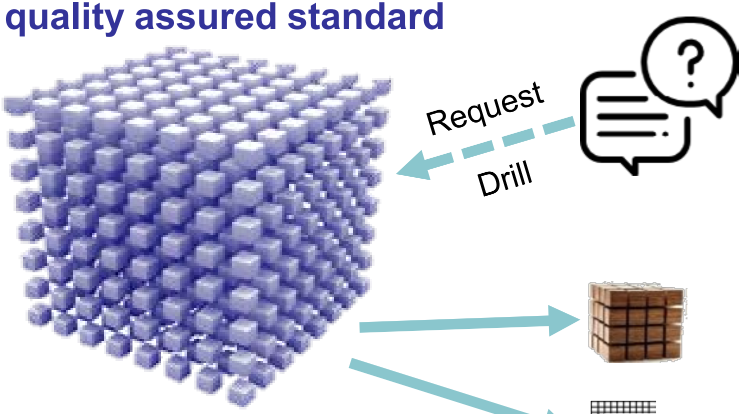
- ➢ Validation at Basic Cube level has positive affects at all other levels
- ➢ OeNB's/ECB's/EBA's DQM reduced
- ➢ Basis for passive data concept
- ➢ Need for machine-to-machine communication

Development of validation rules by

- ➢ **OeNB, AuRep, banks in cooperation**

Passive data concept - the future of reporting?

Basic Cube – quality assured standard

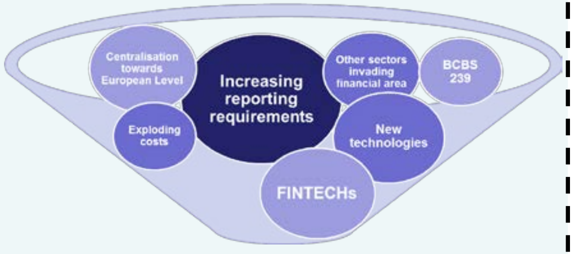


- Specific drills/data requests only when needed (e.g. onsite inspection)
- via standardised language
- e.g. use of APIs

- Reduced set on regular transmissions
- Especially for time series analysis (e.g. economics, financial stability) and key indicators

Conclusions

Challenge



Solution

= Austrian integrated reporting data model

- represents a **paradigm shift** in regulatory and statistical data remittance.
- fosters two-way **understanding** and **transparency.**
- offers new ways in **DQM** and **data analytics**
- Leads to
  - higher **consistency** and **data quality**,
  - **less redundant** data deliveries,
  - higher **flexibility**, and
  - expected **lower costs**.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Collect once, use multiple times – the Reserve Bank of New Zealand's data collection approach[1]

Steffi Schuster and Neil Humphries,

Reserve Bank of New Zealand

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Collect once, use multiple times

Reserve Bank of New Zealand's data collection approach

Steffi Schuster & Neil Humphries, Data & Statistics Department, Reserve Bank of New Zealand[1]

September 2019

Abstract

Like many, the Reserve Bank of New Zealand (RBNZ) derives banking sector statistics from detailed surveys. In recent years, the Statistics team at the RBNZ has developed new surveys to collect prudential and statistical data from registered banks. The new surveys filled data gaps and improved the consistency of our data. What has changed, however, is how we have gone about this.

In this paper we are discussing the new banking sector data collections to illustrate our approach and vision for collecting data.

The RBNZ's paradigm is to collect data once, use multiple times. This means a centrally managed approach to data collection where on a subject matter (e.g. registered banks' balance sheets) a single data collection can satisfy the needs of supervisors, analysts and statisticians.

Before introducing any changes to our data collections we put significant effort into understanding the questions our internal and external data users are seeking answers to and keep them engaged throughout the process. In our experience, working collaboratively with respondents results in significant data quality gains.

We aim to make sure all data and concepts are consistent within and across surveys. Our detailed bank balance sheet survey is at the centre and all other surveys can be linked back to this collection; the concept of a hub.

Some of the metrics from individual banks are now published. In May 2018 the RBNZ introduced the Bank Financial Strength Dashboard – an easy to use online tool for comparing all New Zealand incorporated banks. In total, 110 metrics are available for each bank.

We are reflecting on our vision for data at RBNZ and looking for opportunities to become more responsive to the demands of data users while considering the burden on respondents.

[1] Unless otherwise stated, views expressed are those of the authors, and do not necessarily represent the views of the Reserve Bank of New Zealand.

Contents

Filling data gaps - in recent years the RBNZ has been on a journey…

Our experience at the Reserve Bank of New Zealand (RBNZ), particularly since the global financial crisis (2008), had been as new policy questions arose demand for data increased. Initial attempts to react quickly and fill some of data gaps were only partial and did not align well to best practise. Over the last five years we have changed our approach. This is our story.

The mandate and customers

As a background, the RBNZ is a relatively small, "full service" central bank with 274 staff[2]. In addition to traditional central bank functions we are the prudential regulator of the financial system.

Our main objective is to promote a sound and dynamic monetary and financial system and to ultimately raise New Zealand's economic wellbeing.

We undertake a range of activities to meet this objective, including:

- formulating and implementing monetary and financial policy
- licensing and prudential supervision of banks, non-bank deposit takers and insurers
- supply and circulation of currency
- operating New Zealand's payments and settlement system

RBNZ is policy maker, regulator, and also has a guardian function. As a result we have a need to collect both prudential and statistical data. This function is a centralised service provided by the RBNZ's Data & Statistics department. We aim to collect data that meets the needs of all users, enabling consistent and coherent analysis for different purposes. Our approach to 'collect once, use multiple times' recognises the burden data collection poses for reporting entities and the need to have a single answer in support of enduring policy questions.

Expectations of data users, the example of registered banks data

Table 1

Data user	Use case and expectations
Prudential supervision	Accurate and timely read of individual supervised entity. Ability to compare entities.
Statistician (official statistics)	Accurate and timely read of entities. Data quality is fit for aggregation and use in statistics compiled in alignment with international standards.
Registered bank providing data	Only relevant data is collected and used for monitoring and decision making. Accurate and timely published data (mainly aggregates) for market share and other business performance analysis.
Financial analysts, economists, etc.	Accurate and timely read of the financial system as a whole with ability to identify drivers for change, e.g. change in funding or credit provision.

[2] Reserve Bank of New Zealand Annual Report 2018-19 https://www.rbnz.govt.nz/about-us/annual-reports/2019-annual-report (accessed 30/09/19).

The intent of the prudential data collection is ultimately to allow individual entities, peer groups and overall system presentation of the data.

Published statistics usually present system aggregated data following internationally agreed standards (e.g. the IMFs Monetary and Financial Statistics Manual). The focus of these statistics is more on financial instruments and (counterparty) institutional sectors. RBNZ is a provider of official statistics and the data collected by RBNZ also informs official statistics, e.g. macro-economic statistics provided by Stats NZ.

Our approach gives users confidence that there is only "one source of the truth" for each supervised entity and even enables a consistent and transparent disclosure regime (RBNZ's Bank Financial Strength Dashboard presentation - more on this later).

Most of the regular data collections run by RBNZ concern the banking sector. In terms of supervision and prudential regulation historically RBNZ's focus had been on banks. The Insurance (Prudential Supervision) Act 2010 introduced an insurer licensing regime and RBNZ became the supervisor and regulator of the licensed insurers. Non-bank deposit takers are regulated by RBNZ but not supervised. (See Appendix A for an overview of New Zealand's financial system).

Unlike the setup of some central banks and financial sector regulators in other jurisdictions the RBNZ mandate includes monetary and financial policy, and prudential supervision of regulated entities (e.g. banks and insurers) – a wide range of objectives which opens up the need and opportunity to centrally coordinate statistical and prudential data collection to avoid duplication of effort and unnecessary burden on data providers. This arrangement is not very common yet in other jurisdictions but recognised as desirable.

About RBNZ data collections

Who do we collect data from?

RBNZ uses its legal authority to regularly collect data from all regulated entities and other financial institutions relevant for financial stability and/or monetary policy. This includes: registered banks, non-bank deposit takers, licensed insurers, fund managers, non-deposit-takers, nominees, and registries.

What do we collect?

To date most of RBNZ's regular data collections are "macro" in nature, and usually monthly or quarterly reports to be provided in a pre-defined MS Excel file. RBNZ mainly seeks summarised data. Most data reporting templates require respondents to aggregate up customer data, i.e. to sum by product or into 'brackets' of loan value. Over time the granularity of data requested has increased but as a general rule RBNZ does not request individual customer data. There are however some exceptions (issuers of securities and bank large or connected credit exposures information). All data supplied to RBNZ is protected and not made available to third parties unless authorised.

The data reporting templates are available on the RBNZ website: https://www.rbnz.govt.nz/statistics/surveys.

Data (quality) expectations

Data providers 'own' their data. RBNZ expects the data submissions to be accurate and no adjustments are made to the data of individual entities to derive summary statistics (e.g. system total). If any validation or data quality issues are identified, revised data submissions are required from the data provider.

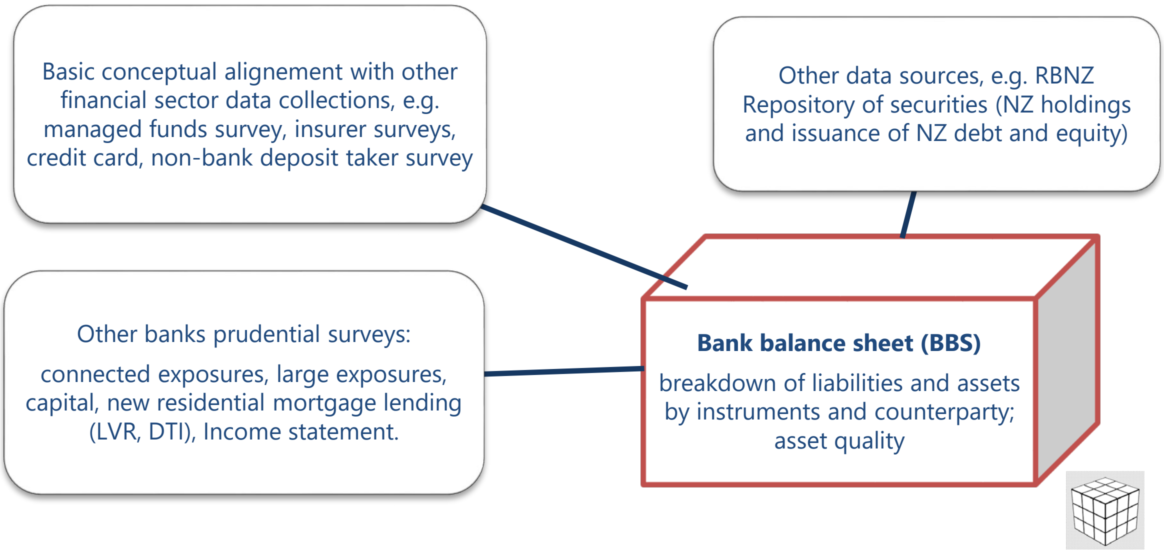
Drivers for change and our vision

In the past – particularly immediately following the global financial crisis (2008) - instead of taking a centralised approach, various teams at RBNZ have drafted (overly simplified) new data requests for regulated entities (mainly banks) to complete. This approach resulted in data quality challenges because different collections had some inconsistent definitions, making it difficult and costly for entities to report and for RBNZ analysts difficult to quality assure. A lot of time and effort was spent on reconciling across surveys to build up a coherent picture of the financial sector.

As the number of regular data collections continued to increase we changed our approach. In this paper we discuss how all recently introduced and redeveloped data collections are designed to achieve consistency across the different dimensions of the financial sector.

We illustrate the example of the redeveloped (registered) bank balance sheet data collection:

Figure 1: A simplified version of the hub concept



In 2015 we started to develop the bank balance sheet (BBS) data collection replacing the previous statistical balance sheet collection from registered banks, the Standard Statistical Return (SSR).

The intent of the new BBS collection was to create a single high level balance sheet which is the central "hub" for the related data collection suite, with the majority of other collections effectively satellites that link back into some or many of the balance sheet's components. The BBS is more granular than the previous SSR. This has enabled much of the linking across surveys. A key outcome is to be able to collect data once but use it multiple times for supervision, macro-prudential and statistical purposes.

The BBS project has helped us move closer to our vision. We are reviewing existing data collections and their alignment with the hub concept. We apply this concept to all new data collections. However for some data collections full harmonisation back to the BBS will be difficult. For example, some compliance concepts cannot be fully harmonised across all surveys given the very specific policy need.

Looking ahead we are also considering the benefits and suitability of collecting even more granular data, which may reduce burden for data providers and better meet the needs of data users.

Making it happen - the change process

To avoid the mistakes of the past we agreed on a centralised and iterative development approach, which was led by the Statistics function. Part of the planning included stakeholder analysis. We needed to better understand the questions the various teams were seeking answers to.

Working with internal stakeholders

We connected with internal data users and invited them to join an internal working group throughout the development of the BBS. The group included supervisors, policy analysts and economists.

We started by asking some simple questions about their data needs such as:

- Why do they need it?
- How will they use it?
- When is it required?

The group assessed the types of instruments that may be needed and the types of breakdowns of the instruments (by institutional sectors or counterparties, pricing, products and industry classifications). Drafting would take many iterations. One of the lessons is that one question probably leads to many. Policy analysts do not always know precisely what they want until the information actually arrives, so the development process is often iterative. Similarly, the definitions of data items will develop over time and need refinement as new events or classification challenges arise.

To be able to start and pilot with small sections of data, and checking in to confirm the new information will align to the business drivers, has proved more valuable than trying to deliver the full new collection in one go.

But working with respondents (our key external stakeholders) has produced some of our biggest gains in data quality.

Working with the key external stakeholders

Over the last five years we have built up our relationships with respondents - with the banks in particular, but also with non-banks (insurers, managed funds, etc.). We engaged with reporting analysts, managers and IT staff. We had many one-on-one meetings and ran workshops. Building relationships, providing clarity of our requests and spending time to discuss is not costless but extremely valuable in the long run.

Developments are costly to the provider and the Bank in terms of:

- Staff time, system changes or enhancements;
- Crowding out of competing stakeholder projects; and
- May require culture as well as simple processing changes.

Building relationships, however, opens the opportunity for respondents to "buy-in" or become a stakeholder in the collection, improve their understanding of why the collection is needed and knowledge as to what questions are being addressed. Respondents can gain some ownership in the collection and can (and do) make suggestions to improve the format and definitions to ensure the end result aligns with their own understanding of how the sector operates, aligns with their own management reports which significantly improves their validation processes, and may start to answer questions they may have of their own - particularly in relation to market share. From the RBNZ's perspective, we gained more knowledge on the operational aspects of banking and clarity around business definitions.

The key benefit is the improvement in data quality.

The banks contributed and agreed consistent definitions, after thorough consultation (internal and external workshops). As much as practically possible definitions are aligned with industry accepted terminology and diagrams included to visualise the requirements (see Appendix B – a visual decision tree for loans and advances).

We found visual presentation of the requirements helped gathering feedback and helped reduce misunderstandings significantly. For example, workshop participants may appear to have agreed but have walked away with a different understanding as not all participants have precisely the same assumptions in mind.

Definition inconsistencies are both inefficient and burdensome for all involved in surveys and statistical outputs. Consistency in definitions across collections reduces time in the reconciliation process for both the banks and for the RBNZ. Definitions, however, remain subject to refinement and the need to reflect changes in accounting standards.

Other issues impacting on quality include:

- Data providers trying to add metadata layers to existing systems designed for different purpose;
- Seeking information from multiple systems that do not align; and
- Fitting multiple products into one cell in a collection.

In addition to building relationships, the opportunity to trial new collections as well as ultimately parallel running the new collection has enabled all involved to test and provide feedback leading to collection template and definition improvements. Some issues are not actually identifiable until respondents try to complete the collection - good in theory but practise is the test!

New surveys collect both prudential and statistical data

Some of our new collections are for both prudential and statistical data purposes. The bank balance sheet (BBS) collection, introduced in March 2017, is such an example.

The new BBS collection is a single high level balance sheet which is a central hub for related data collection suite, with the majority of other collections effectively satellites that link back into some or many of the balance sheet's components (see Appendix C). A key outcome is being able to collect data once but use it multiple times for prudential and statistical purposes.

From a prudential perspective it captures the key financial instruments, and counterparties, provisions and net adjustments, as well as asset quality measures for individual banks. For statistical outputs, the focus is more on system or banking sector as a whole reporting, with highlights on specific counterparties such as households, businesses and specific financial instruments (i.e. mortgages, consumer loans and deposits). Individual bank responses are aggregated (summed) to enable publication for the system. Statistical outputs tend to be reported gross, in line with international reporting requirements (IMF). By capturing provisions and net adjustments separately, multiple reporting requirements can be met.

We are now collecting more granular data in the BBS than its predecessor, the SSR. By going more granular in our collection design we achieved the versatility discussed above. Instead of requesting summarised data we are collecting more detail which allows data to be used multiple times in different but related data concepts, e.g. net to gross values, and reconciliation across banking surveys.

Bank Financial Strength Dashboard

Following a regulatory stock take the Bank Financial Strength Dashboard was developed to support the New Zealand bank's disclosure regime[3]. The RBNZ launched the interactive Bank Financial Strength Dashboard in May 2018. In March 2019 the Dashboard won the Central Banking Best Initiative of the Year Award 2018.

The introduction of the BBS, and several other new prudential collections (Capital and Large exposure prudential satellites, Loan-to-value lending positions, Asset quality and an updated Income Statement survey), has enabled consistently defined and timelier data to be available. Data captured in the current Liquidity survey will also be updated over the next couple of years.

Unlike individual banks' disclosure statements which banks publish on their own websites, the Dashboard shows banks' financial information side by side on a comparable basis, and in a central location. The RBNZ updates the Dashboard quarterly, 40 working days after the reporting period, by extracting information from the above surveys submitted by the New Zealand-incorporated banks and their related banks.

The Dashboard features:

- Individual bank data published in an interactive form;
- Banks as contributors and also as users;
- Around 110 metrics sourced from various RBNZ banking data surveys.

It's a Journey that is ongoing…

We have been able to achieve significantly improved regular data collections by working closely with stakeholders. Through this process we have been able to achieve our aim to "collect once and use multiple times" for a significant range of banking sector data. We have reviewed what we need to ensure data and statistics are relevant, fit for purpose, cost efficient and effective both in terms of collection and use.

We want to build on these lessons and lift the bar higher by seeking more granular data, e.g. considering anonymised customer-level transactional data. Most of our data collections are not as detailed and therefore flexible as we would like them to be. Collections are still mainly aggregations of multiple products for some respondents.

We also want to widen the collection tool options and so potentially reduce the need for survey respondents to aggregate data into cells in a MS Excel template. This should reduce the burden on respondents as data remains in its native format as much as possible.

This approach, however, will introduce new challenges. It will increase the effort required by RBNZ as the collecting entity to better understand individual respondent metadata as each bank is likely to have uniquely labelled metadata due to branding and competitive interests in each institution. They built their computer systems to meet their business needs and not our reporting requirements.

[3] Tobias Irrcher; Liam Gillies; Cavan O'Connor-Close The dashboard - year one: shining a brighter light on the financial strength of banks https://www.rbnz.govt.nz/research-and-publications/reserve-bank-bulletin/2019/rbb2019-82-07 (accessed 18/09/2019)

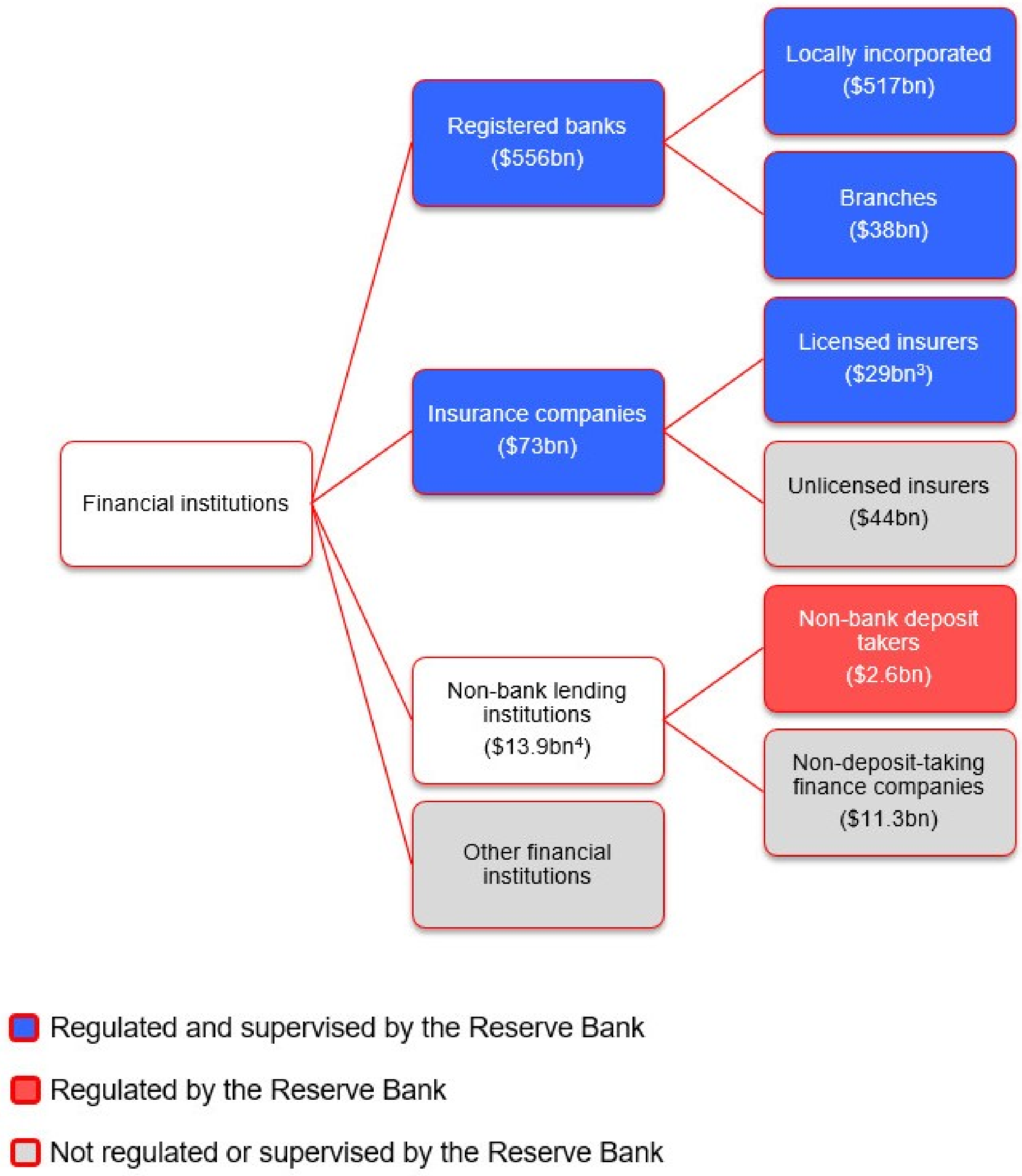
RBNZ already has some experience with granular data. The Repository of Securities captures individual financial instruments and individual legal entities (issuers and holders) data which is used in the regular production of debt instrument statistics.

We will build on our experience.

RBNZ is embarking on a multi-year programme investing in its data systems. As part of the modernisation programme we will investigate the feasibility and merits of transactions data for providers and data users.

Appendix A: Overview of the New Zealand financial system

Financial institutions' total assets (as at 31 March 2019)

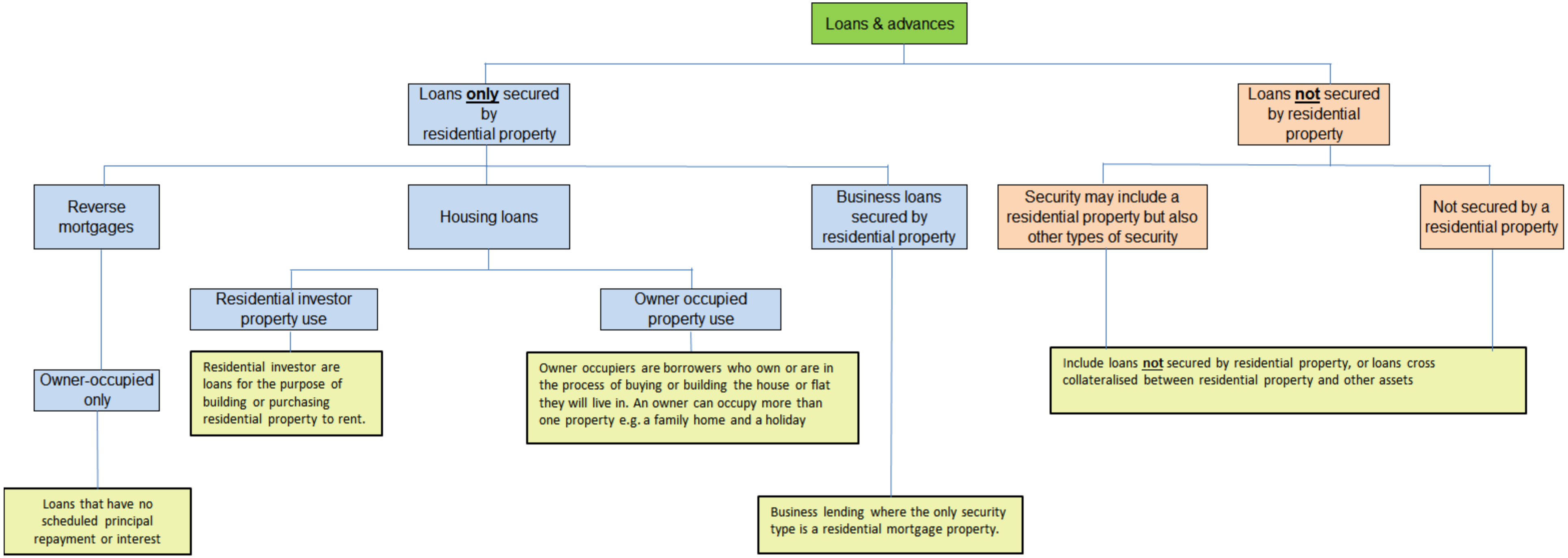


Note: Numbers may not sum due to rounding.

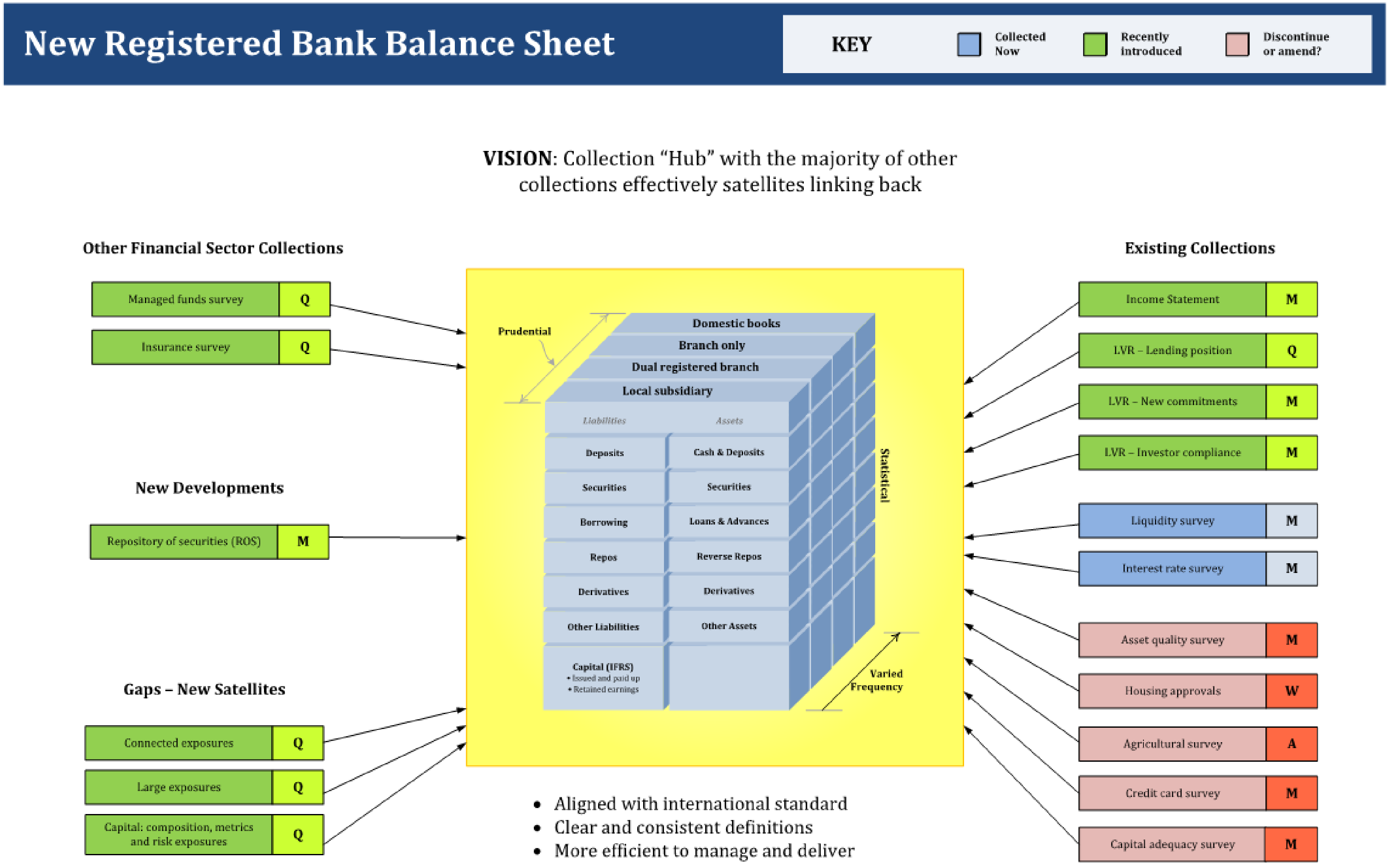
Source: RBNZ Bank Balance Sheet (BBS), RBNZ *Standard Statistical Return* (SSR), RBNZ Non-Bank Deposit Taker Prudential Return (NBDTPR), Individual Insurer Financial Statements

Source: https://www.rbnz.govt.nz/financial-stability/overview-of-the-new-zealand-financial-system (retrieved 4 July 2019)

Appendix B: Loans & advances decision tree



Appendix C: High level vision (as in 2016)



Collect once, use multiple times

RESERVE BANK
OF NEW ZEALAND
TE PŪTEA MATUA

Collect once, use multiple times

The Reserve Bank of New Zealand's data collection approach. (It is a journey...)

Steffi Schuster & Neil Humphries
Reserve Bank of New Zealand
16th August 2019, 9.15 am – 10.30 am

Collect once, use multiple times...

1. About New Zealand's central bank (RBNZ)
2. RBNZ's data collection approach. It is a journey…
    - Single collection for multiple purposes;
    - Working with stakeholders;
    - Workshops, pilots, trials, parallel runs; and
    - Lessons
3. Bank Financial Strength Dashboard
4. The journey is on-going…

RBNZ functions and objectives

The Reserve Bank of New Zealand (RBNZ) main objective is to promote a sound and dynamic monetary and financial system and to ultimately raise New Zealand’s economic wellbeing.

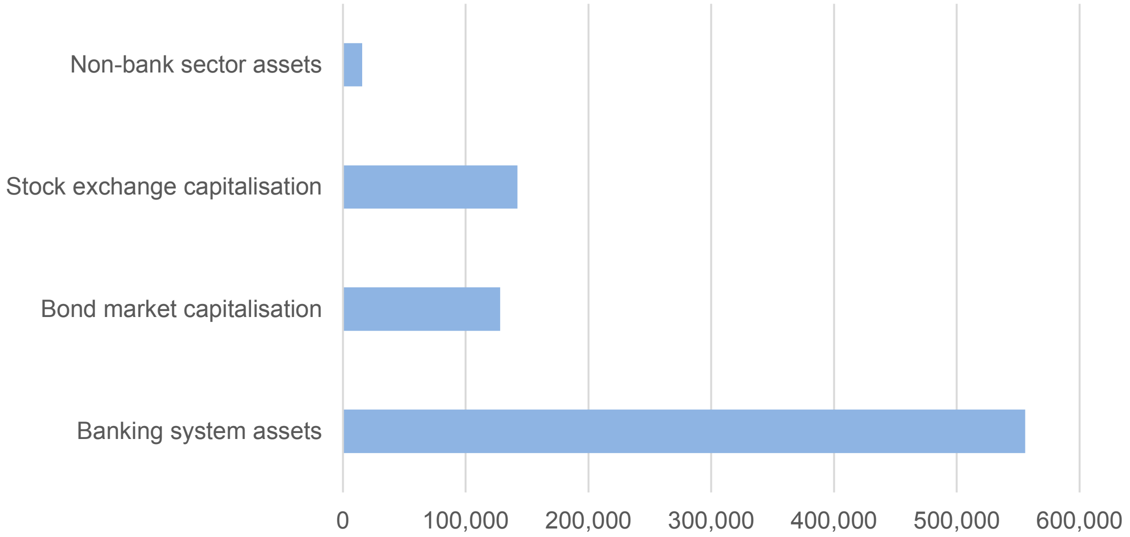
Our activities:

- formulating and implementing monetary and financial policy
- licensing and prudential supervision of banks, insurers and other deposit taking firms
- supply and circulation of currency
- operating New Zealand’s payments and settlement system

RBNZ is policy maker, regulator, and also has a guardian function. As a result we have a need to collect both prudential and statistical data.

NZ’s financial system

Size of New Zealand’s financial sector (March 2019, NZ$ billion)



Source: RBNZ, NZX
Note: Bond market capitalisation excludes private debt placements. Stock exchange capitalisation excludes unlisted share holdings.

Surveys and guiding principles

All regular data collections are:

- Relevant
- Fit for purpose
- Cost effective

RBNZ collects regular data from all regulated entities and other financial institutions relevant for financial stability and/or monetary policy:

- All registered banks, non-bank deposit-takers & licenced insurers file prudential data returns
- Funds manager (e.g. superannuation), non-deposit taking non-banks, etc. provide statistical returns

We expect data to be accurate and will not make adjustments. If changes are needed the data provider will have to resubmit.

- ➢ Consistency of aggregates and individual entities reported data (e.g. enables accurate market share analysis)

We work with data providers when (re)designing data collections.

Stock-take and new approach

As data demands increased (post 2008) not all data collections were aligned, causing inefficiencies.

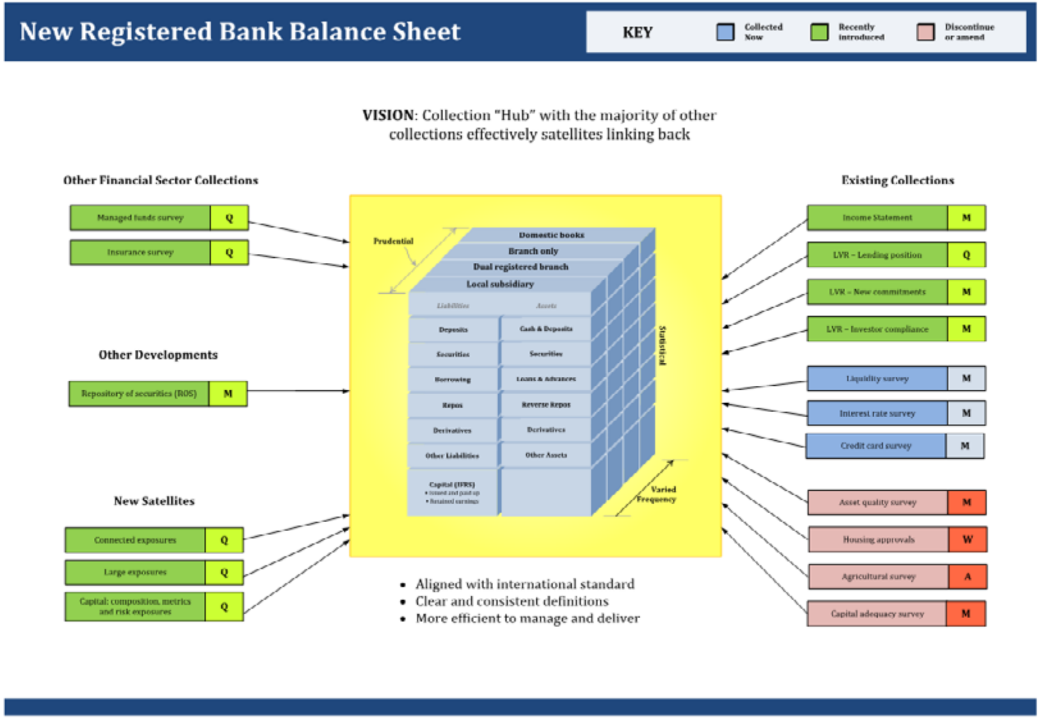
<u>Our approach last 5 years</u>: All recently introduced and redeveloped data collections are designed to achieve consistency across the different dimensions of the financial sector.

<u>Example</u>: redeveloped (registered) bank balance sheet data collection

<u>The vision</u>: the bank balance sheet (stock positions at the end of each month) is a hub and the majority of other data collections can be linked to it.

Single collection for multiple use…

Prudential, Macro-prudential & Statistical purposes (Gross & Net)



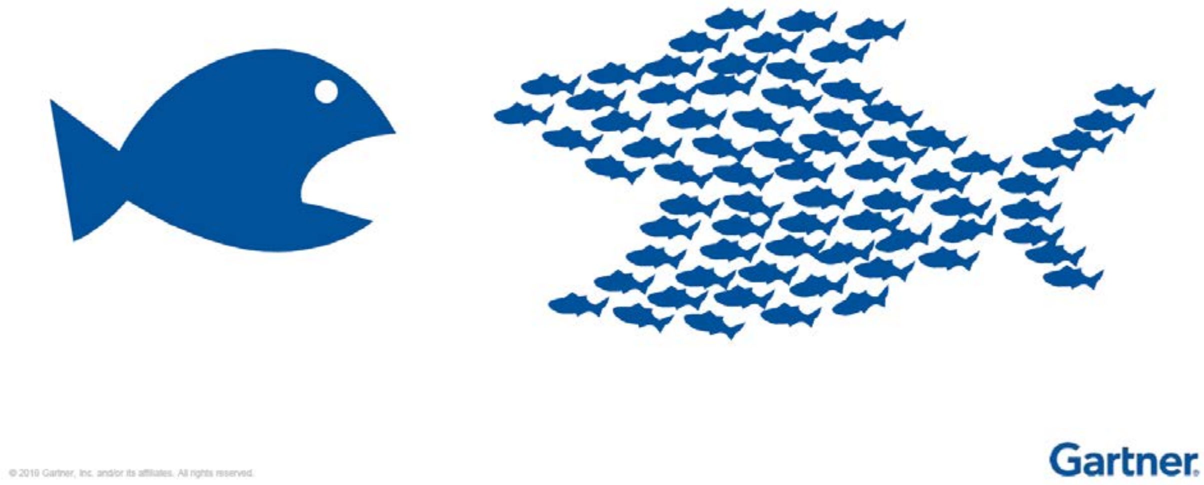
As at 2017

Work with stakeholders…

Internal
- Better understand questions that policy makers want answers to…
- Take users on a journey…
- Why, How, When? One question can lead to many…

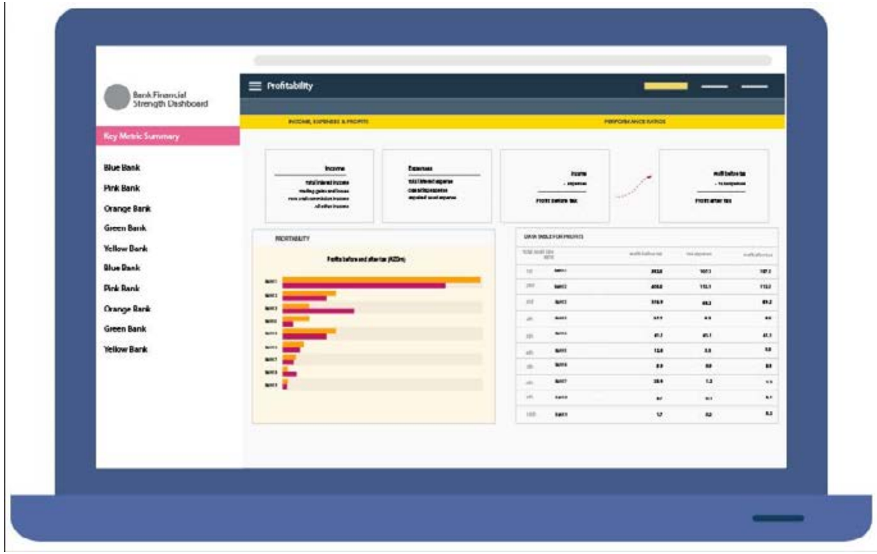
External
- Relationship building, one-on-one, peer group & workshops
- "Buy-in" and "ownership" i.e. industry agreed and consistent definitions
  → data quality improvements

Workshops, Pilots, Trials, Parallel runs…



Bank Financial Strength Dashboard

In March 2019 RBNZ received the Central Banking Best ‘Initiative of the Year’ award for the [Bank Financial Strength Dashboard]().



The journey is ongoing…

All regular data collections are:
- Relevant
- Fit for purpose
- Cost effective

Next…
- Modernising all data collections (end to end)
- Reduce level of aggregation
- Explore use cases for loan-level data

More challenges…
- Reduce burden on respondents
- Increase effort to understand and align meta data
- Evolution of user needs (e.g. supervision and policy analysis)

Thank you

Steffi Schuster & Neil Humphries

Reserve Bank of New Zealand

16th August 2019, 9.15 am – 10.30 am

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Best practice recommendations on facilitating access to microdata: outcomes from the INEXDA working group on data access[1]

Stefan Bender,

Deutsche Bundesbank

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Best practise recommendations on facilitating access to microdata: Outcomes from the INEXDA working group on data access

Stefan Bender
Deutsche Bundesbank

IFC-BNM-ECB Satellite Seminar on
"Post-crisis data landscape: micro data for the macro world"
August 2019

Motivation

- **Aggregate datasets** are important for **monitoring macroeconomic developments** and **macroeconomic policy**
- **Granular data** is necessary to understand **global developments** and in particular **differences across countries**
- Combining datasets and looking beyond aggregate statistics into heterogeneous developments require the **transformation** of **“data“** into **“knowledge“**
- **Local constraints** make it difficult, or often impossible, to link micro datasets from different jurisdictions, even for research and financial stability analysis
- **Better accessibility** and **sharing of granular data** would open up **new possibilities** for analysis by providing new **insights into the effect of policies**

What can **we do** from the **statistical side** to support this process?

INEXDA: The Granular Data Network

On 6th January 2017,

have launched the International Network of Exchanging Experiences on Statistical Handling of Granular Data (INEXDA), an international cooperative project to declare their willingness to further strengthen their cooperation.

Since its foundation, the following institutions have joined INEXDA as members:

Chronology of past INEXDA meetings

1st INEXDA meeting in Lisbon

- INEXDA members (DE, FR, IT, PT, UK)
- *Guests:* BIS

2nd INEXDA meeting in London

- INEXDA members
- *Guests:* BIS, ECB, ES

3rd INEXDA meeting in Paris

- INEXDA members (+ECB, ES)
- *Guests:* AT, BIS, CL, MX, TR, UK (NSI)

4th INEXDA meeting in Basel

- INEXDA members (+CL, TR)
- *Guests:* AT, CH, BIS, DE (NSI), Eurostat, MX, RU, UK (NSI)

5th INEXDA meeting in Frankfurt

- INEXDA members (+Eurostat, Russia)
- *Guests:* AT, BIS, DE (NSI), FR (NSI), GR, NO (NSI), UK (NSI)

Jan 2017 — Jul 2017 — Jan 2018 — Aug 2018 — Apr 2019

Jan 2017:

INEXDA
The Granular Data Network

Memorandum of Understanding

Signing and publication

Jul 2017:

da|ra

INEXDA Metadata Tool

by GESIS

Jan 2018:

Working groups

1. Dissemination
2. **Metadata**
3. **ADRF**
4. Modes of accreditation
5. Contracts for research projects/bodies
6. Modes of data provision
7. Output control
8. Risk management for published results

Aug 2018:

Data Access

Apr 2019:

Workshop on Data Access

Working Group on data access

1 INEXDA workshop on data access procedures

- Serve as a kick-off event for the INEXDA work stream on data access
- Utilise current momentum to improve data access procedures for data producers and researchers.

2 Define common access procedures

- Identify similar workflows used by data providers.
- From there define a standard
- Need to be in compliance with internal data policies and external government regulations.

3 Development of open software solution

- Set of common access procedures will define software requirements
- Needs to be flexible enough to accommodate different legal frameworks

INEXDA Workshop on access procedures *(April 2019)*

1 **Data discovery**: providing information on where to find the data

- Ivana Ilijasic Versic, Consortium of European Social Science Data Archives (CESSDA)
- Ines Drefs, GO FAIR
- Hendrik Doll, Deutsche Bundesbank

2 **Data access centres:** how to grant access to the data: nationally and internationally

- Paulo Guimarães, Banco de Portugal
- Kamel Gadouche, Centre d'accès sécurisé aux données (CASD)
- Beate Lichtwardt, UK Data Archiv (UKDA)

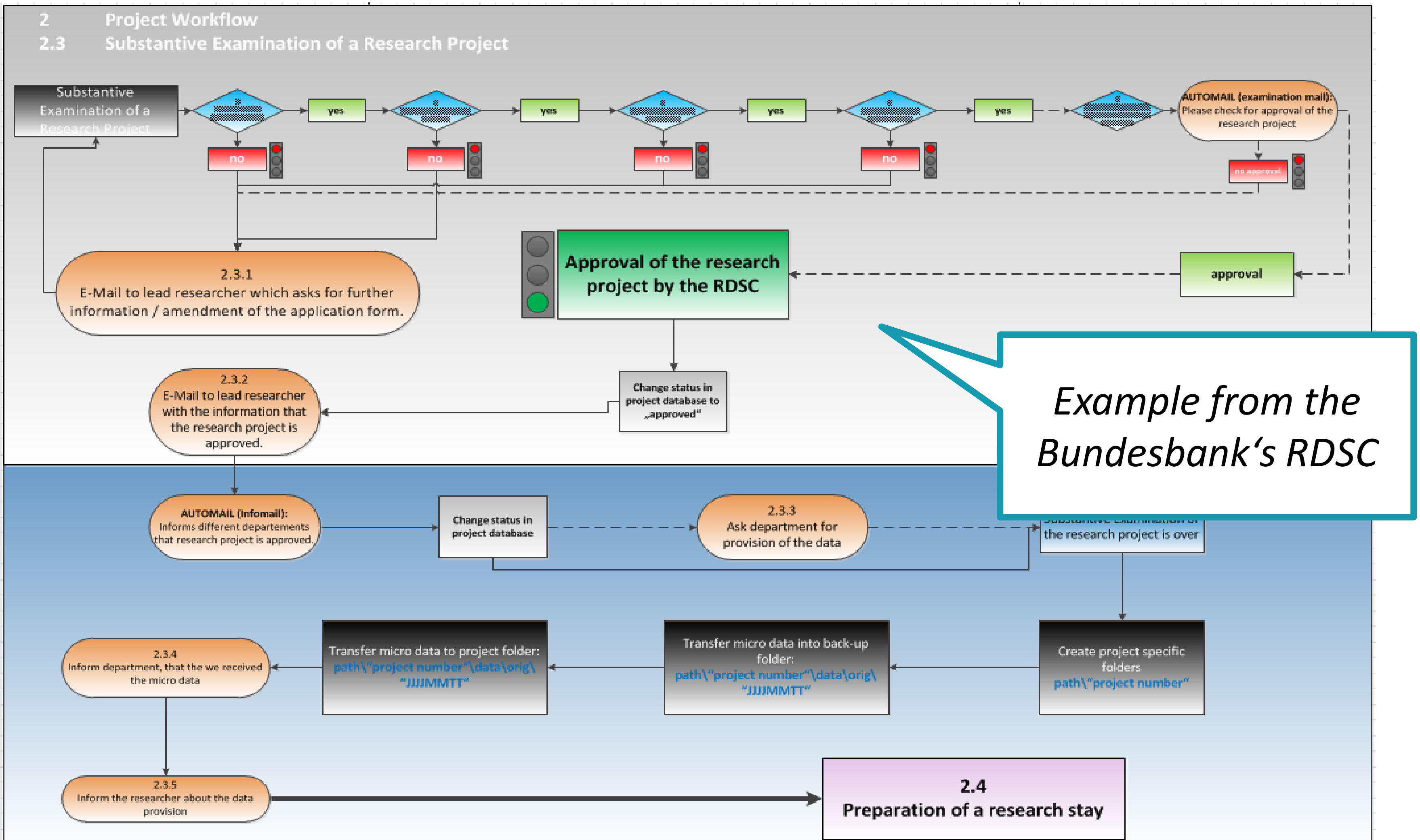
3 **Alternative access models for multi-source data**

- Marina Randriamisaina, Banque de France
- Fabio Ricciato, Eurostat
- Johan Heldal, Statistik Sentralbyra (SSB)

4 **International Initiatives on micro data sharing**

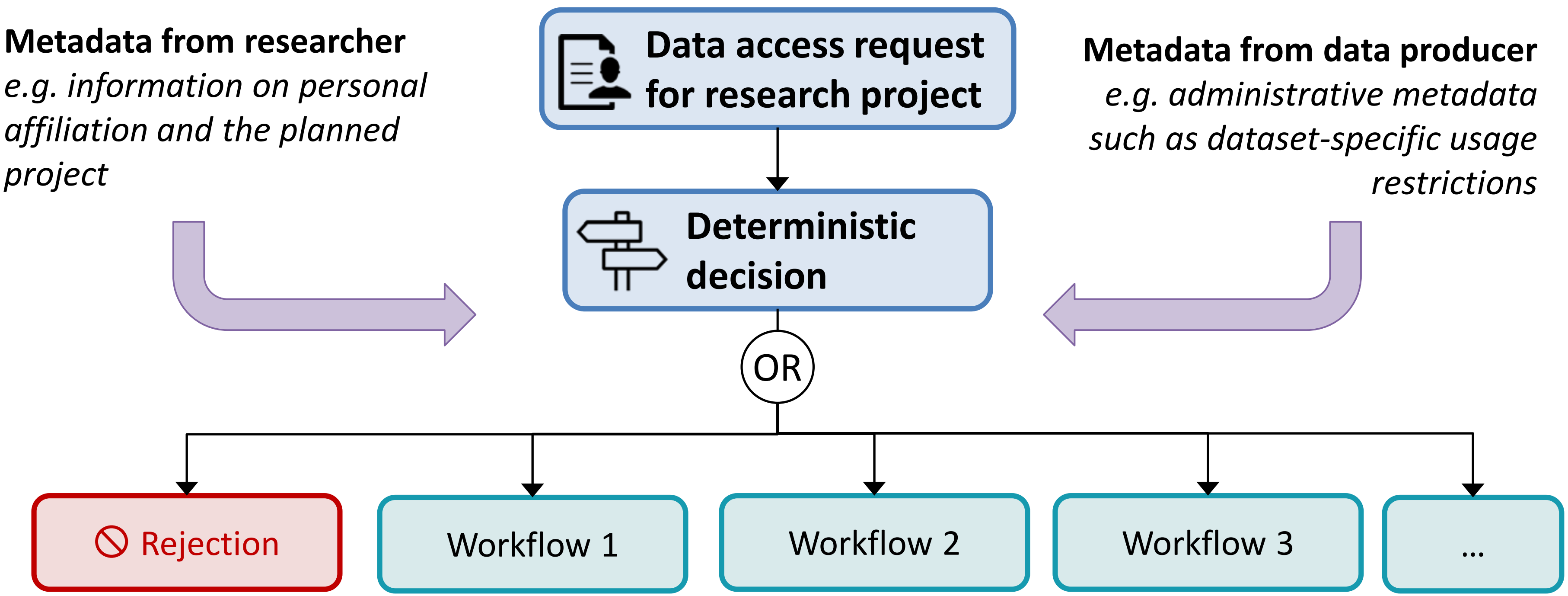
- Emily Witt, European Central Bank (ECB)
- Roxane Silberman, International Data Access Network (IDAN)
- Claus Goran Hjelm, Nordic Microdata Access Network (NordMAN)

Status quo: Very complex to standardise multiple workflows



Ideal state: Automation of decisions in workflows

- ✓ **Metadata** describes how data can be managed and used from a legal, organisational or technical point of view
- ✓ **Workflows follow deterministically** from such administrative dataset metadata



Conclusion

- INEXDA provides a platform for exchanging experiences on statistical handling of granular data for central banks, national statistical institutes and international organizations.
- INEXDA supports the G20 process, especially the Data Gaps Initiative 2 recommendation aiming to promote the exchange of (granular) data as well as metadata.
- Focus of the INEXDA Data Access Working Group is to define common access procedures and use these for the development of an open software solution.

Next steps

End of August 2019	Collection of feedback on proposed new metadata items
September 2019	Meeting of the INEXDA Data Access Working Group
End of 2019	Proof of concept for open source data stewardship module software in collaboration with NYU's ADRF

INEXDA

The Granular Data Network

Thank you for your attention!

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Enhancing user knowledge by providing (richer) context to microdata[1]

Christian Hirsch,

Deutsche Bundesbank

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Enhancing user knowledge by providing (richer) context to microdata

Christian Hirsch, Deutsche Bundesbank

IFC-BNM-ECB Satellite Seminar on
"Post-crisis data landscape: micro data for the macro world"
August 2019

Based on a project with and contributions from:
Stefan Bender, Hendrik Doll, Julia-Katharina Ginz, Christian Resch, John Chase, Jonathan Morgan, Ian Mulvaney, Andrew Gordon and Julia Lane

Motivation

Status Quo: One-directional information flow

- Researchers use microdata and metadata to produce outcome (publications)
- No structured knowledge flows back to data producers

Goal: Enable feedback loops

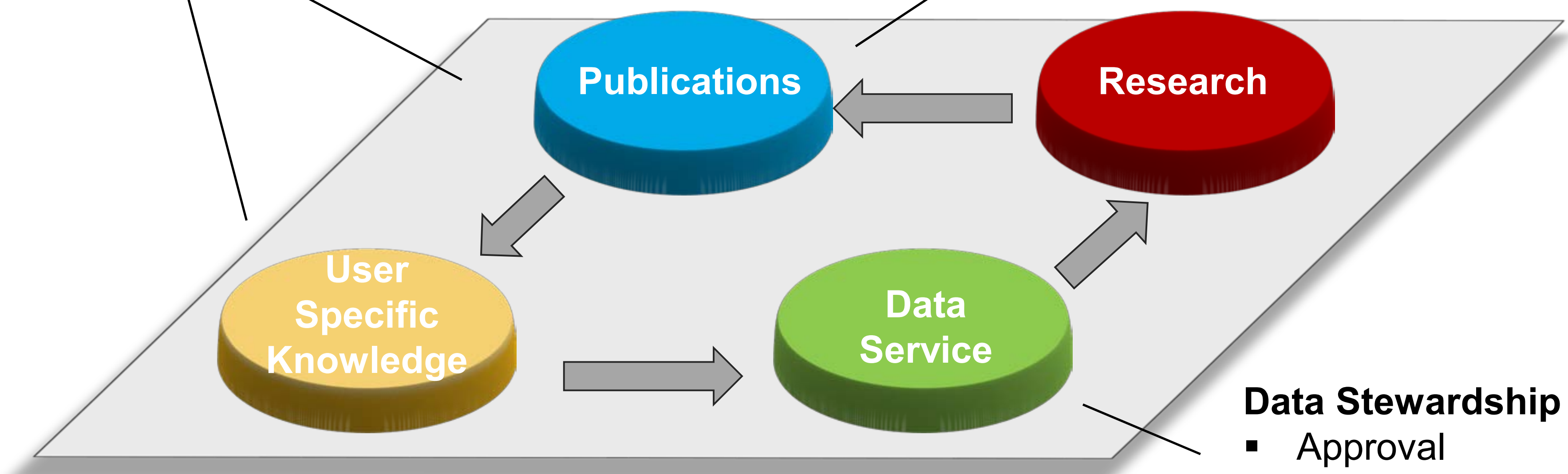
- Make knowledge generated in the research process usable
- Two sources of information to extract:
  1. Automatic: Text mining to find information in research publications
  2. Tacit knowledge: Enable and incentivize data users to feedback information

Knowledge Life Cycle (IDIS-R)II

Collaboration
- Knowledge sharing
- Metadata

Secure workspace
- Services and Tools

Rich Context

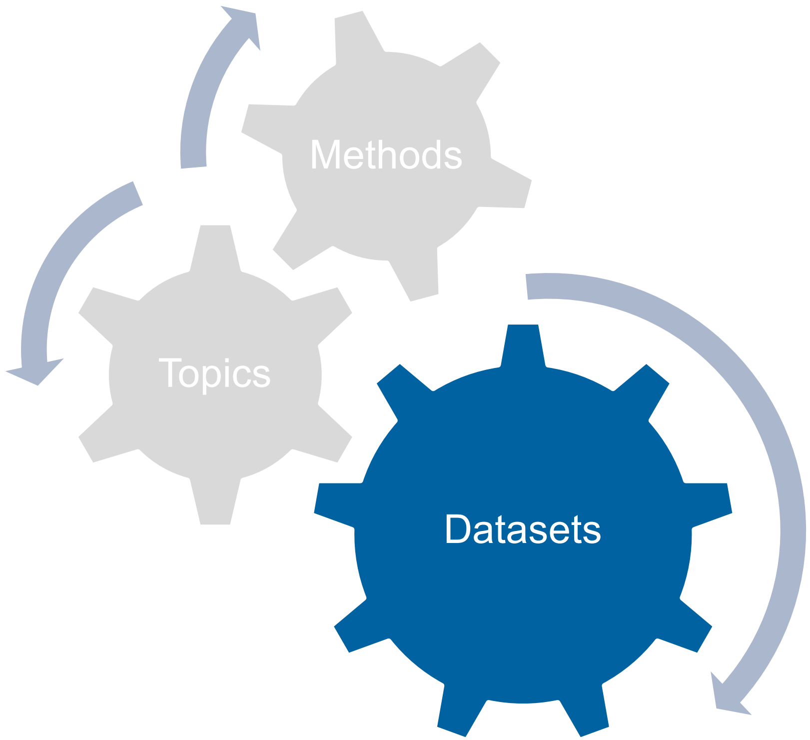


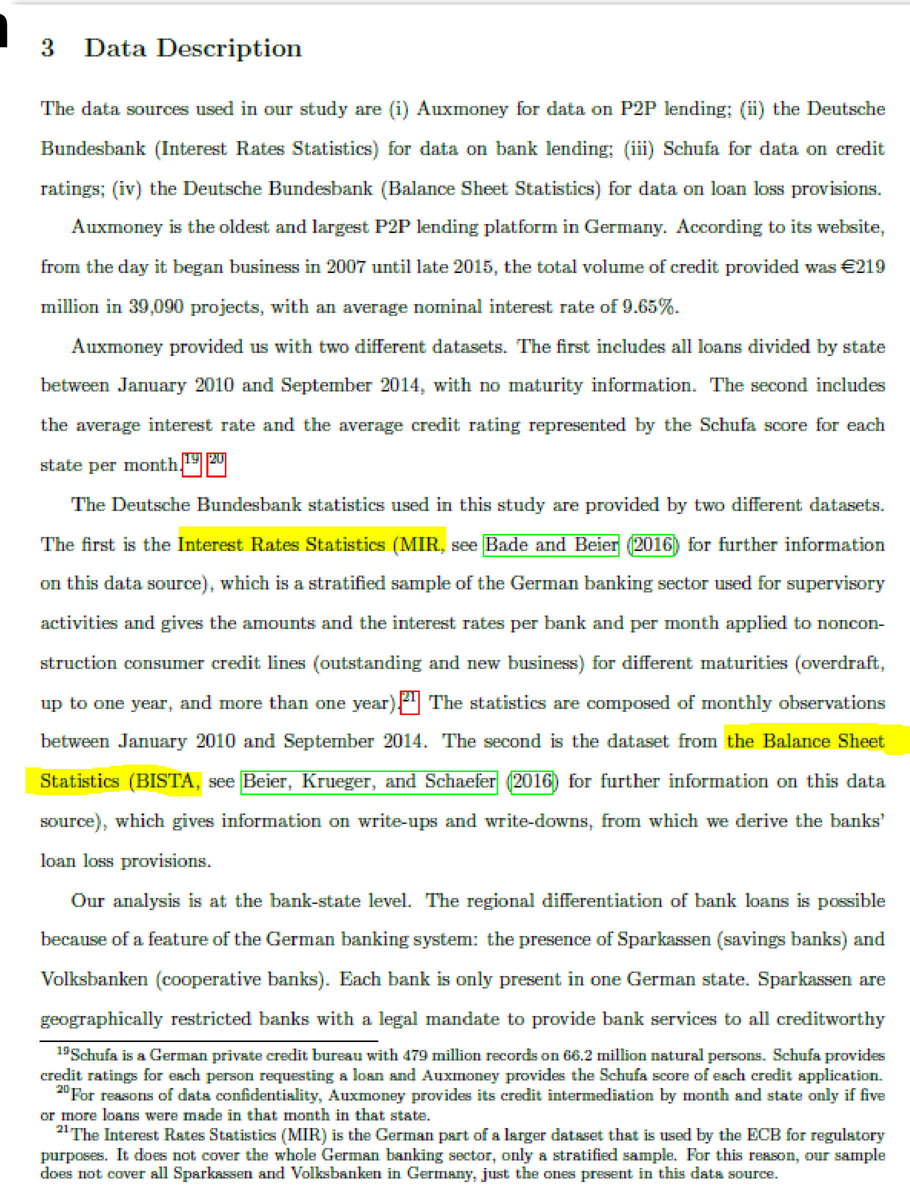
Data Stewardship
- Approval
- Monitoring
- Reporting

How do we do this?

- **Step 1:** Create the set of corpora and usage side metadata (machine learning techniques) - Competition
- **Step 2:** Figure out how you learn from it and automate it (machine learning techniques) – Dataset recommendation engine
- **Step 3:** Engagement – recognize and emphasize patterns (with human curation) – Rinse and repeat

Step 1: competition design





3 Data Description

The data sources used in our study are (i) Auxmoney for data on P2P lending; (ii) the Deutsche Bundesbank (Interest Rates Statistics) for data on bank lending; (iii) Schufa for data on credit ratings; (iv) the Deutsche Bundesbank (Balance Sheet Statistics) for data on loan loss provisions.

Auxmoney is the oldest and largest P2P lending platform in Germany. According to its website, from the day it began business in 2007 until late 2015, the total volume of credit provided was €219 million in 39,090 projects, with an average nominal interest rate of 9.65%.

Auxmoney provided us with two different datasets. The first includes all loans divided by state between January 2010 and September 2014, with no maturity information. The second includes the average interest rate and the average credit rating represented by the Schufa score for each state per month.[19] [20]

The Deutsche Bundesbank statistics used in this study are provided by two different datasets. The first is the Interest Rates Statistics (MIR, see Bade and Beier (2016) for further information on this data source), which is a stratified sample of the German banking sector used for supervisory activities and gives the amounts and the interest rates per bank and per month applied to nonconstruction consumer credit lines (outstanding and new business) for different maturities (overdraft, up to one year, and more than one year).[21] The statistics are composed of monthly observations between January 2010 and September 2014. The second is the dataset from the Balance Sheet Statistics (BISTA, see Beier, Krueger, and Schaefer (2016) for further information on this data source), which gives information on write-ups and write-downs, from which we derive the banks' loan loss provisions.

Our analysis is at the bank-state level. The regional differentiation of bank loans is possible because of a feature of the German banking system: the presence of Sparkassen (savings banks) and Volksbanken (cooperative banks). Each bank is only present in one German state. Sparkassen are geographically restricted banks with a legal mandate to provide bank services to all creditworthy

[19] Schufa is a German private credit bureau with 479 million records on 66.2 million natural persons. Schufa provides credit ratings for each person requesting a loan and Auxmoney provides the Schufa score of each credit application.

[20] For reasons of data confidentiality, Auxmoney provides its credit intermediation by month and state only if five or more loans were made in that month in that state.

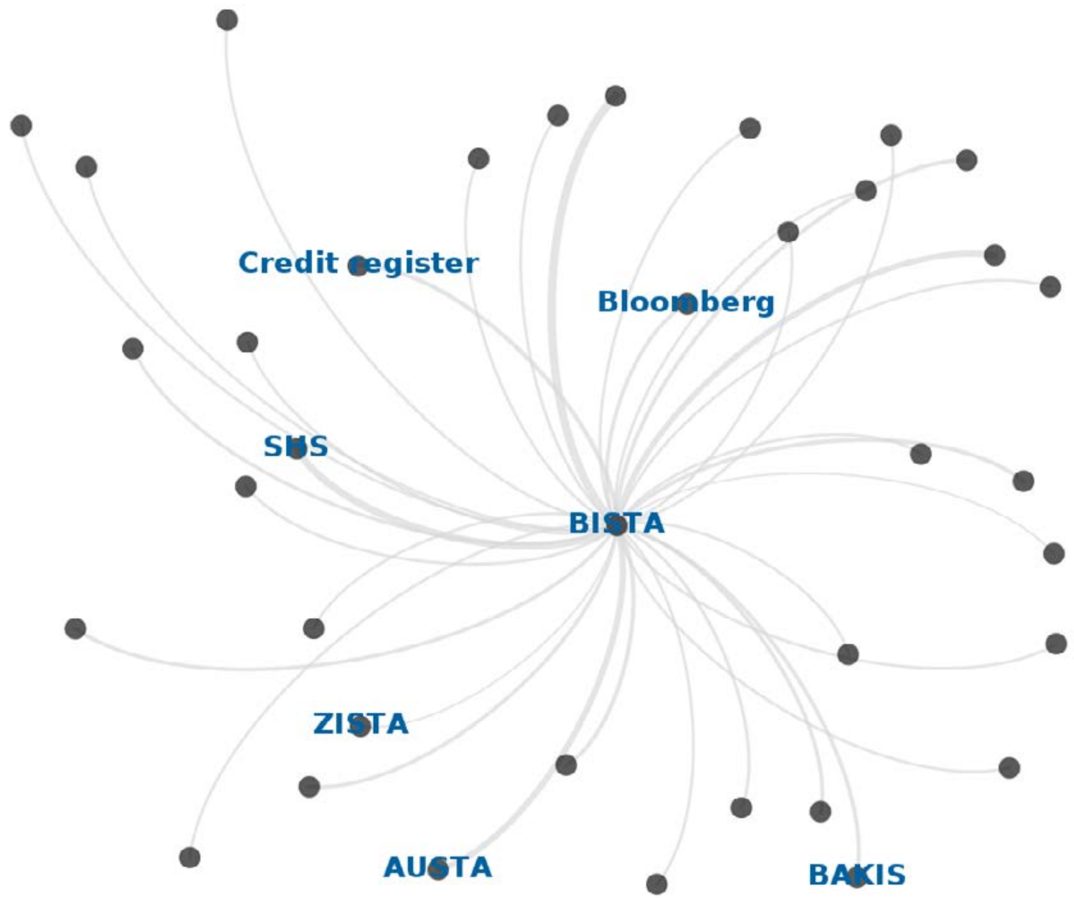
[21] The Interest Rates Statistics (MIR) is the German part of a larger dataset that is used by the ECB for regulatory purposes. It does not cover the whole German banking sector, only a stratified sample. For this reason, our sample does not cover all Sparkassen and Volksbanken in Germany, just the ones present in this data source.

Calebe de Roure – Loriana Pelizzon – Anjan Thakor

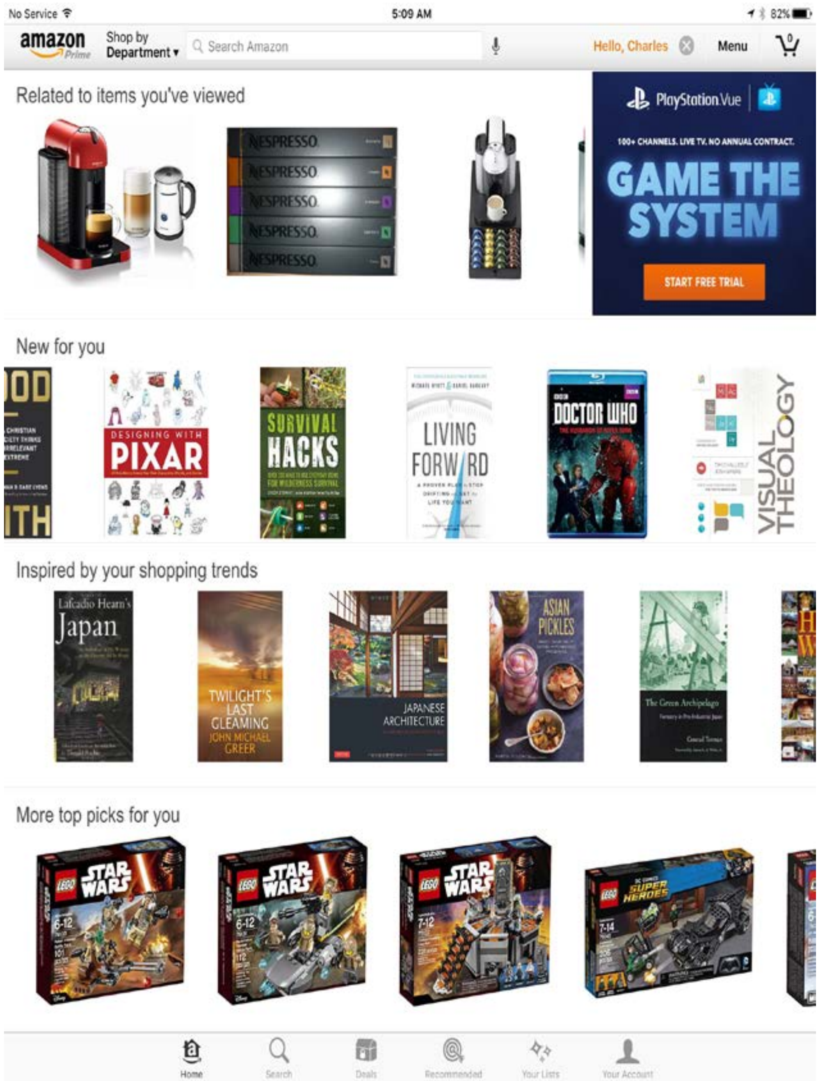
P2P Lenders versus Banks:
Cream Skimming or Bottom Fishing?

SAFE Working Paper No. 206

Step 2: use of user specific knowledge - input



Step 2: use of user specific knowledge



Related to data you've viewed

New data similar to data you've used

What others have done with similar data (recipes)

Recipes like yours

Thanks to Julia Lane

Interaction with INEXDA's metadata schema

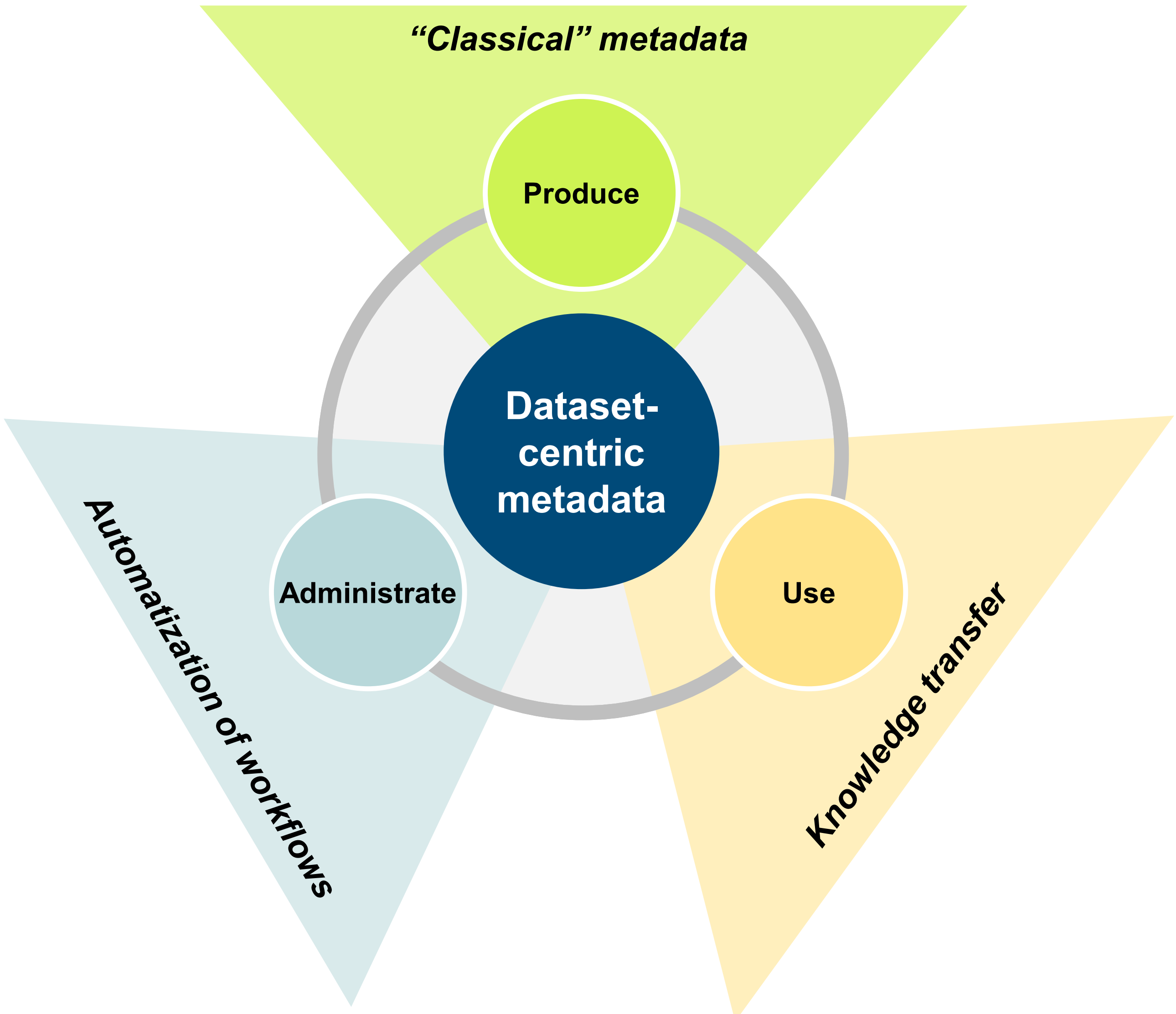
1	Resource Type
2	Resource Identifier
3	Name of Dataset
4	Creator
5	DOI Proposal
6	URL
7	Language of Resource
8	Publication Date
9	Availability
10	Sampled Universe
11	Sampling
12	Temporal Coverage
13	Time Dimension
14	Collection Mode
15	Unit Descriptions
16	Descriptions
17	Geographical Coverage
18	Keywords
19	Alternative Identifiers
20	Relations
21	Publications

- Based on the GESIS DOI registration service da|ra (GESIS is cooperating with DataCite):

  https://www.da-ra.de/en/home

- Is basis for INEXDA metadata database that was established to store and view metadata from INEXDA members.
- Conclusions:
  - Items are helpful for producing meaningful dataset recommendations for researchers (e.g. Temporal Coverage)
  - Need to amend schema to include usage metadata generated from project

Amend existing metadata to include information on use



Conclusion

WHAT WE DID

- Assess feasibility of meaningful dataset recommendations for researchers based on dataset recognition in publications
- Design a mechanism enabling automated knowledge flow back from publications
- Develop a concept to engage users to contribute knowledge to existing metadata

WHAT'S NEXT

End of Aug 2019	Pilot of dataset recommendation tool
15-16 Nov 2019	Rich Context Workshop hosted by NYU
2020	On-board tool to DSM (if possible)

Contact

Christian.Hirsch@bundesbank.de

Website:
https://www.bundesbank.de/en/bundesbank/research/rdsc
Contact: fdsz@bundesbank.de

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Post-crisis skills landscape: the emergence of "purple people"[1]

Luis Teles Dias,

Bank of Portugal

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

BANCO DE PORTUGAL
EUROSYSTEM

Post-crisis skills landscape: the emergence of "purple people"

IFC-BNM-ECB Satellite Seminar "Post-crisis data landscape: Micro data for the macro world"

Kuala Lumpur, 16 August 2019

Luís Teles Dias

Deputy Director | Statistics Department

In numerous central banks MICRODATA is becoming the nucleus of the statistical function.

But in recent years things have gone much far…

THE NEW DATA LANDSCAPE OF THE POST-CRISIS HAS PERMEATED CENTRAL BANKS IN MANY WAYS

- Microdata have become more granular
- Projects involving big data are more frequent
- The interest of central banks in big data is evolving from a mere interest area of research to an auxiliary input (or core input) for policy making and supervisory processes
- Very big structured datasets coexisting with very large unstructured data
- Integrated reporting schemes addressing multiple purposes within the central bank are already in place and will continue to be developed
- New sources supplementing (or partially replacing) established reports are becoming commonplace – commercial data, administrative data, social media datasets

THE NEW DATA LANDSCAPE OF THE POST-CRISIS HAS PERMEATED CENTRAL BANKS IN MANY WAYS

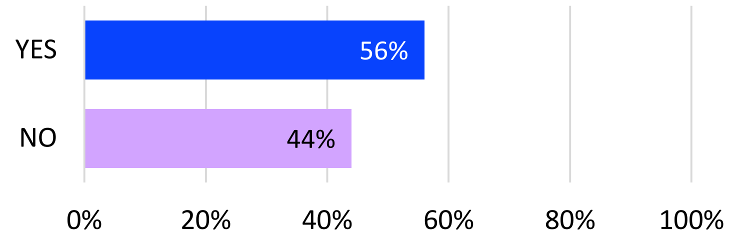
- New approaches to analyse and explore the new data landscape – data mining, visual analytics, machine learning, pattern recognition, etc.

- New data architectures are being implemented to address the need to:
    - **Combine data** from multiple datasets
    - Promote **data-sharing** throughout the institution
    - Allow the **regular** data exploration and also **experimental** activities with the data
    - Increase the **usefulness** of the impressive wealth of data now available not only to the internal users but also, e.g., to the research community

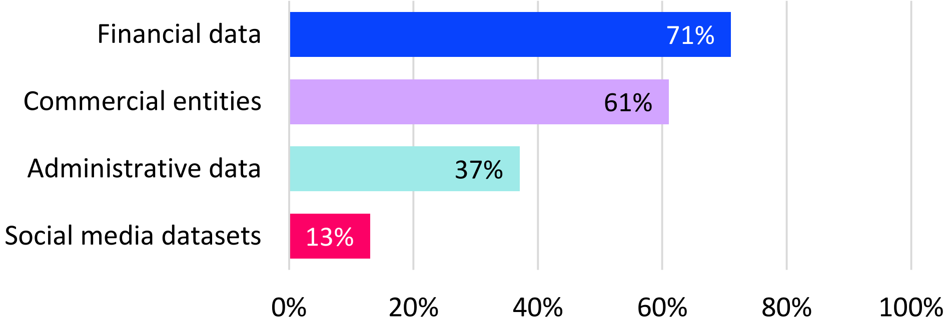
SOME OF THESE TENDENCIES CAN BE QUANTIFIED

“Work in big data can now be considered a mainstream activity for central banks”

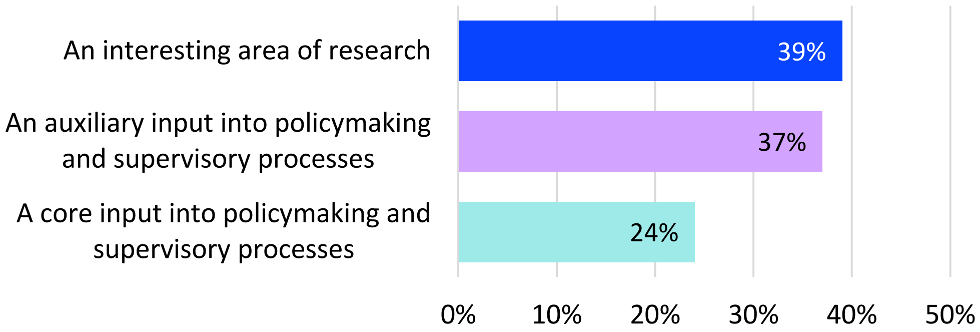
Central banks working on a project involving big data



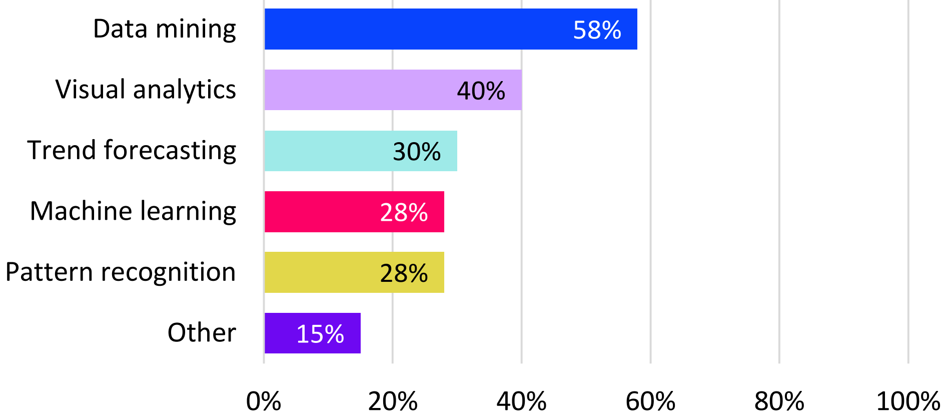
External sources used to obtain big data



Central bank’s view of big data



Approaches used to analyse big data



Source: Findings of the 2018 survey on “Big data and Central Banks” by Central Banking

https://www.centralbanking.com/central-banks/economics/data/3661931/big-data-in-central-banks-2018-survey-results

TWO BIG CHALLENGES

These trends prompt for significant investments in IT infrastructures and software. However what we have experienced at Banco de Portugal is that the greatest challenges stand in **two particular facets**:

DATA GOVERNANCE

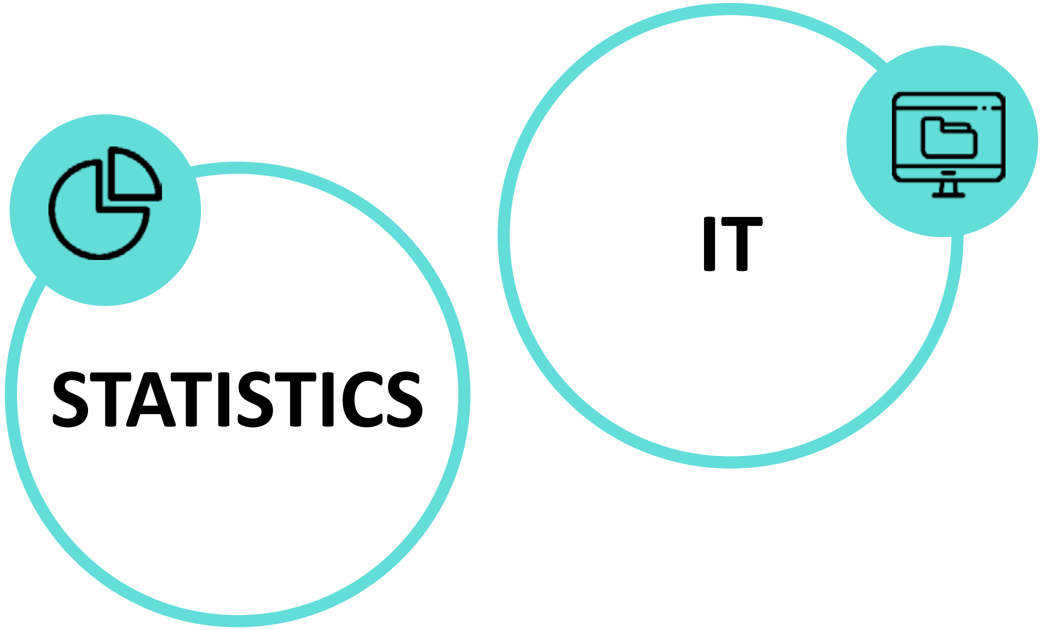
STAFF SKILLS AND TRAINING

In 2017 Banco de Portugal launched the INTEGRATED DATA MANAGEMENT (IDM) programme

The IDM is a major transformational initiative of the Strategic Plan 2017-2020.

The goal of the IDM is to strongly contribute to a better use of the available data in the Bank by means of rationalisation of the processes associated with its collection and processing and to <u>promote its effective sharing throughout the whole organisation</u>.

The IDM is **jointly coordinated** by 2 departments:



SUCCESS FACTORS FOR THE IMPLEMENTATION OF THE IDM

Strong **sponsorship** from the **Board**

Dedicated teams (both in **Statistics** and **IT**)

All departments must recognize themselves in the programme and should **be involved in the decisions**

The level of understanding of the programme by the various departments is not homogeneous. **Adequate expectations management is vital**

Pursuing the global vision supported by **iterative objectives**

Setting an enterprise-wide **DATA GOVERNANCE**

THE MAJOR CHANGE INDUCED BY THE IDM IS NOT TECHNOLOGICAL BUT CULTURAL/ORGANISATIONAL

1st challenge: THE GOVERNANCE MODEL

Centralized coordination with decentralized roles and responsibilities across the organisation

DECISION MAKING [Board]

STRATEGIC COORDINATION [Information and Technology Management Committee]

Information security

OPERATIONAL COORDINATION

Data management
[Statistics Department]

Information technologies management
[IT Department]

DEPARTMENTS

Data stewards	Data owners	System owners
Data experts	Data managers	Data custodians (IT Dep.)

Risk

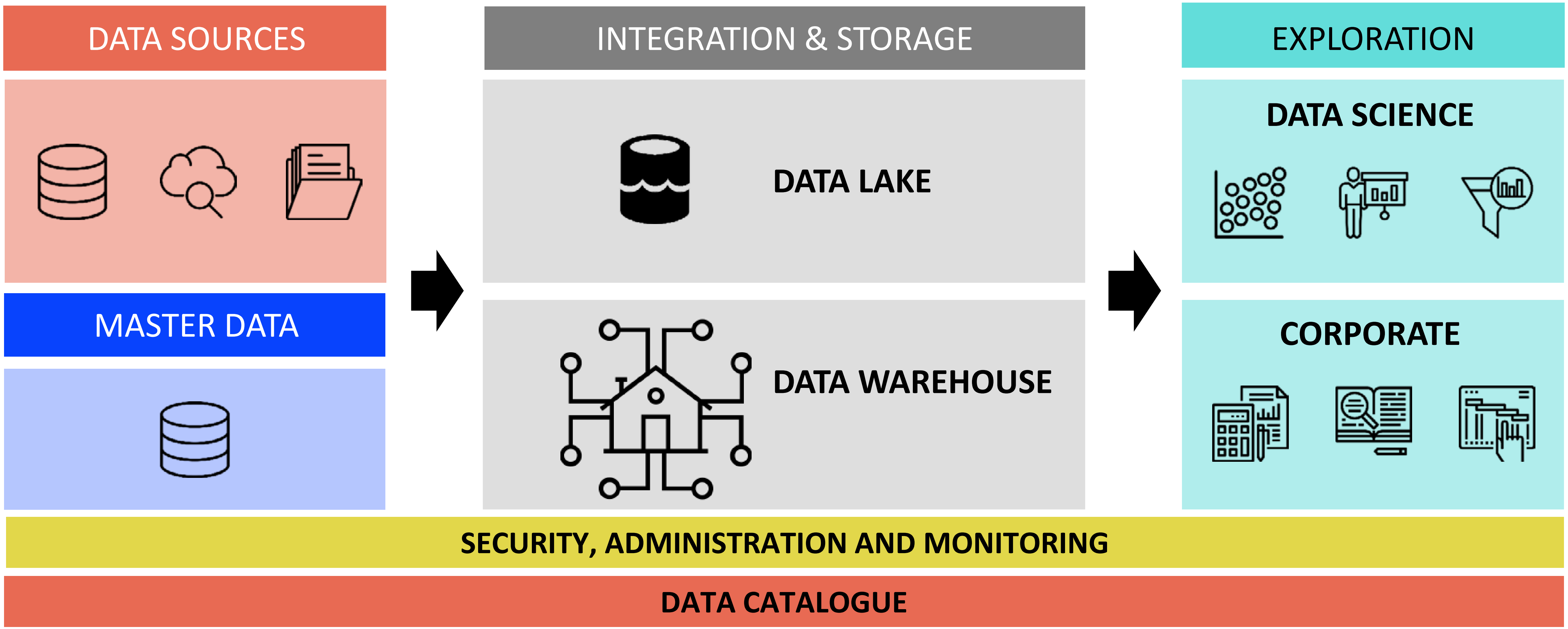
Data protection

Audit

Data providers

Data users

THE LOGICAL DATA ARCHITECTURE



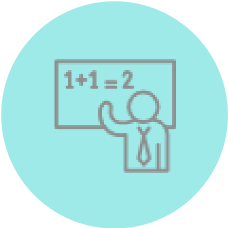
2nd challenge: STAFF SKILLS

ECONOMIC

FINANCE KNOWLEDGE

Business competencies are required.

These are key to perceive what is the relevant information and potential relationship between data sources.



MATHEMATICS

With the increased volume and data variety, mathematical knowledge is needed to be able to perceive correlations and other knowledge that may be hidden in all the data now available.

COMPUTER SCIENCE

Data still need cleaning, transformation, aggregation, etc.

New skills are also required to process unstructured data.

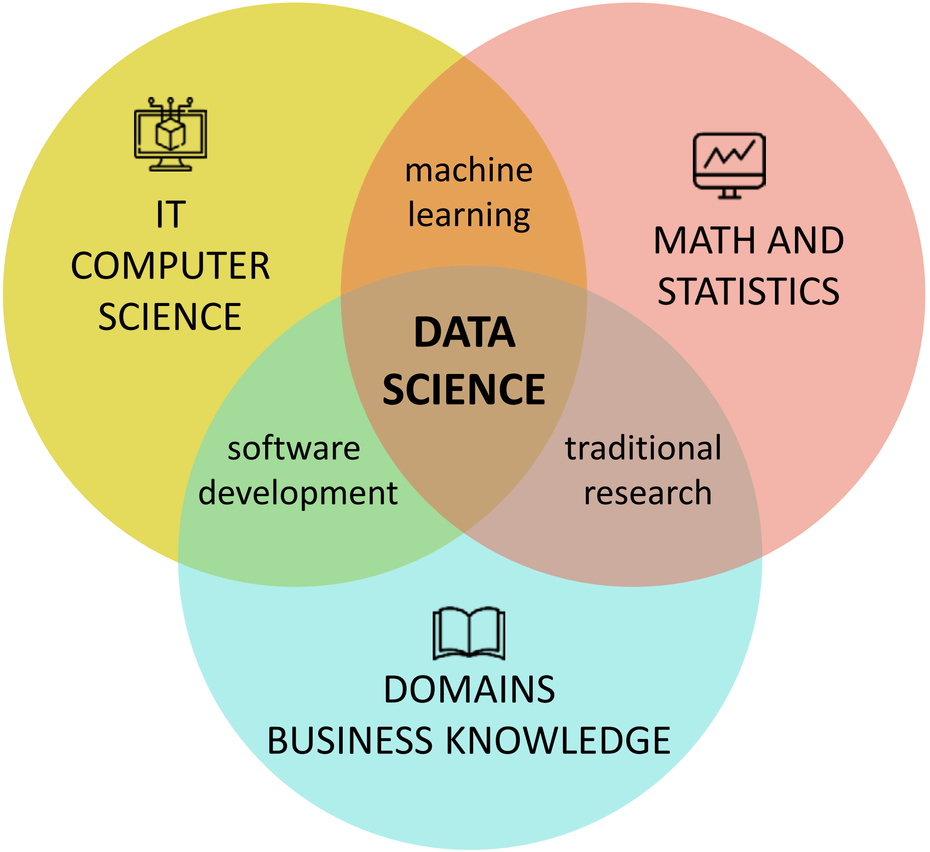
THESE PROFILES ARE IMPORTANT ON THEIR OWN, BUT ARE <u>MUCH MORE VALUABLE WHEN COMBINED</u>

DESPITE NUMEROUS INITIATIVES (road shows in universities, open days, recruitment announcements, …)

ATTRACTING DATA SCIENTISTS TO THE BANK PROVED TO BE AN UNSUCCESSFUL TASK

- To respond to this unsuccess Banco de Portugal decided to launch its own internal training programme – the **SCHOOL OF DATA SCIENCE**

- The Board approved the programme at end-2018 to start in 2020 covering 3 thematic areas



GOAL

To offer a structured and continuous training program aiming at providing the staff in business areas with a blend of skills – TECHNICAL, ANALYTICAL AND COMMUNICATION

THEMATIC AREAS

MATH AND STATISTICS

Data analysis
- Data for decision-making
- Calculations and problems
- Statistics for decision-making
- Data exploration
- Complex analyses

Advanced analytics
- Statistics inference
- Regression analyses
- Predictive analyses and machine learning
- Neural networks and deep learning
- Data mining
- Big data analytics

Data visualisation
- On-screen data models
- Dashboarding
- Storytelling
- Infographics

(…)

DOMAINS BUSINESS KNOWLEDGE

Information management
- Data catalogue
- Master and reference data
- The data warehouse
- Data governance

(…)

COMPUTER SCIENCE / IT

Databases
- Databases concepts
- Introduction to SQL
- Data extraction, transformation and loading
- Data warehouses development

Big data
- Modelling massive amounts of data
- Big data tools

Methodologies
- Agile
- Rapid application development (RAD)
- Business event analysis and modeling (BEAM)
- Parallel Computation
- Distributed Computing

(…)

TECHNOLOGIES and TOOLS

DATA EXPLORATION

- Excel (VBA and excel for BI)
- Tableau
- SAS
- Reporting Services
- R Notebooks
- Jupyter

(...)

ADVANCED ANALYTICS

- Power BI
- Tableau
- SAS
- R
- Python
- Stata
- Matlab

(...)

DATA VISUALIZATION

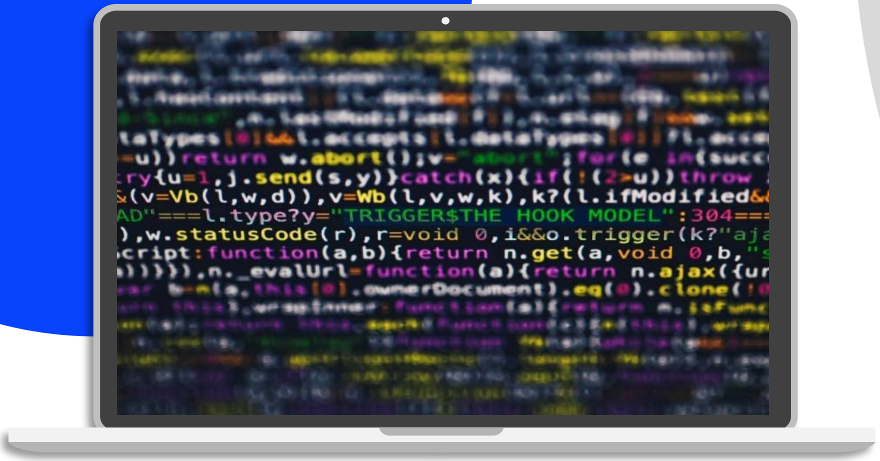
- Excel Power View
- Power BI
- Tableau
- PowerPoint
- D3.js

(...)

DATABASES

- MS SQL Server
- Hadoop ecosystem
- Column store databases

(...)



DIFFERENT PROFICIENCY LEVELS AND AN INCENTIVE FOR EACH STAFF MEMBER TO SHAPE ITS OWN TRAINING JOURNEY

ADVANCED
Strong specialisation in a thematic area

INTERMEDIATE
Content required for a proficient level of professional users

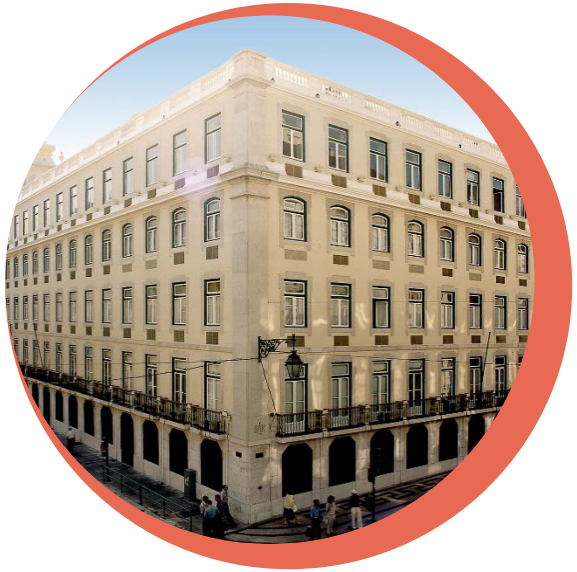
BASIC
Introduction to the different thematic areas

GENERIC
Relevant transversal concepts linked with the data organization in the Bank

LEVERAGE KNOWLEDGE WHEREVER IT EXISTS ...

UNIVERSITIES

TRAINING COMPANIES



BANCO DE PORTUGAL STAFF

- Strategic partnership with a University to assist managing the programme
- Start with a MVP (minimum viable product) for 100 staff members and a subset of 12 courses (including on-line courses) from the thematic areas and from technologies/tools

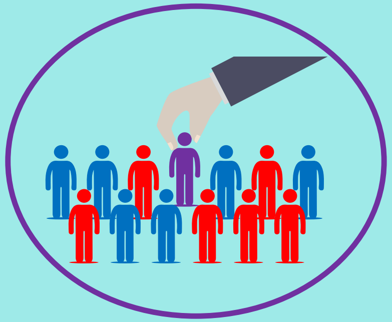
CONCLUDING REMARKS

- Extracting the most from the new post-crisis data landscape requires **organisational transformations**. **Data governance [DG]** and adequate **staff skills** were the major challenges that we have identified at Banco de Portugal

- Besides clear decision making, a successful DG strategy implies a **solid partnership between business and technology**

- DG policies will lose their value if they're not followed in day-to-day operations across the organisation. **Data stewardship** (in all business areas) is now looked as a critical function for the success of the DG

- In a fast-changing technological landscape, **PURPLE PEOPLE**, who can combine data savviness with domain-specific business knowledge, are highly demanded



- In Statistics (but also in other business areas) the challenge is to move from a traditional "economist-statistician" profile to a "economist-data scientist" profile, thus valuing a virtuous **blend** of economic/business training and advanced technological expertise

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Financial behaviour of households under the lenses of distributional aspects: the OeNB's experiences[1]

Johannes Turner,

Central Bank of the Republic of Austria

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

OESTERREICHISCHE NATIONALBANK
EUROSYSTEM

Financial behaviour of households under the lenes of distributional aspects: OeNB’s experiences

Johannes Turner
Director Statistics Department

IFC Satellite meetings
Kuala Lumpur, August 2019

Background: Basic concepts of linking micro and macro

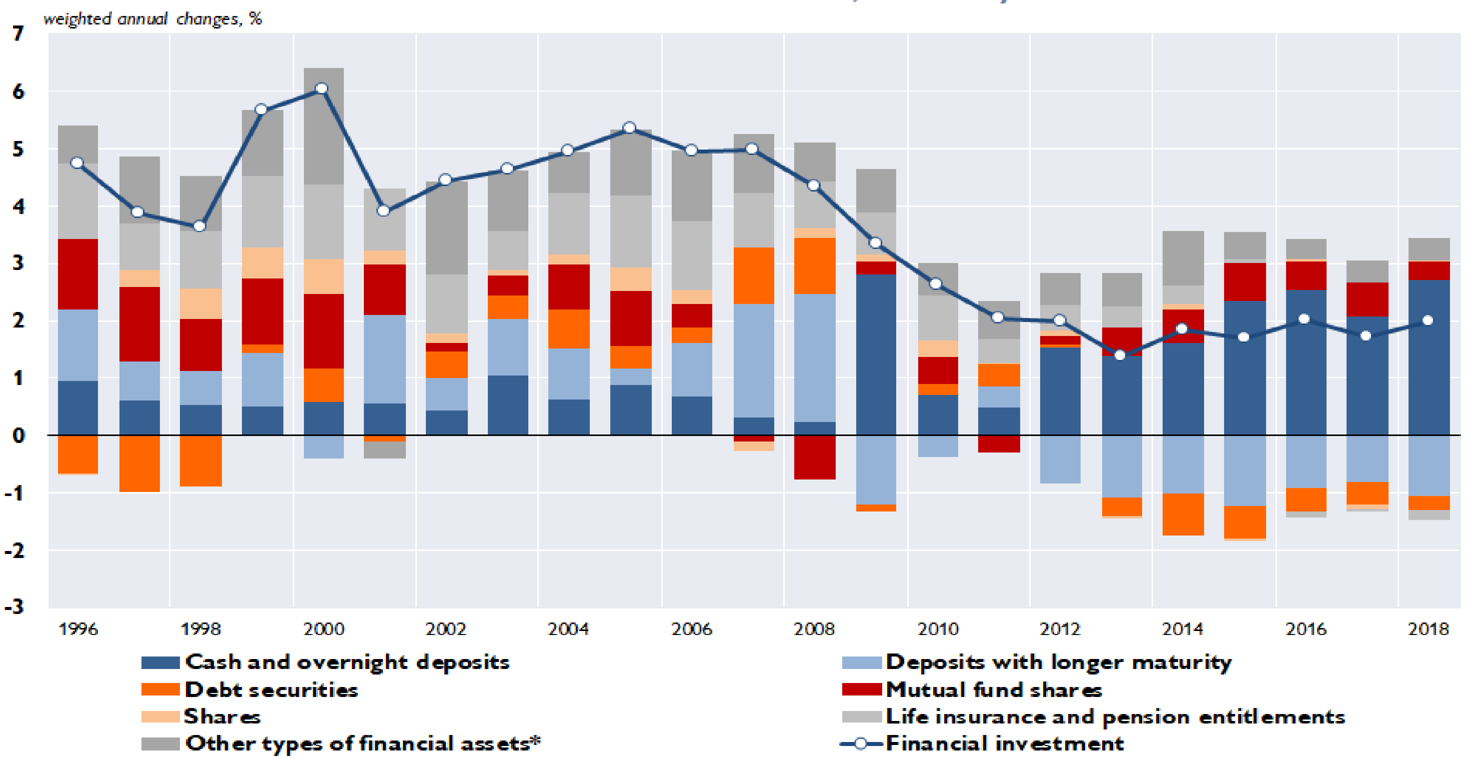
- Drill-down of macro statistics to micro data, aggregated in one statistical framework (using security by security data for BoP)

- Drill-down of different aggregates to micro data using a common data model (using loan by loan data for macro-economic statistics via loans in supervisory statistics)

- **Enrichment of aggregates by socio-economic characteristics (Financial behaviour using information derived from household surveys**)

Recent development in the financial behaviour of households

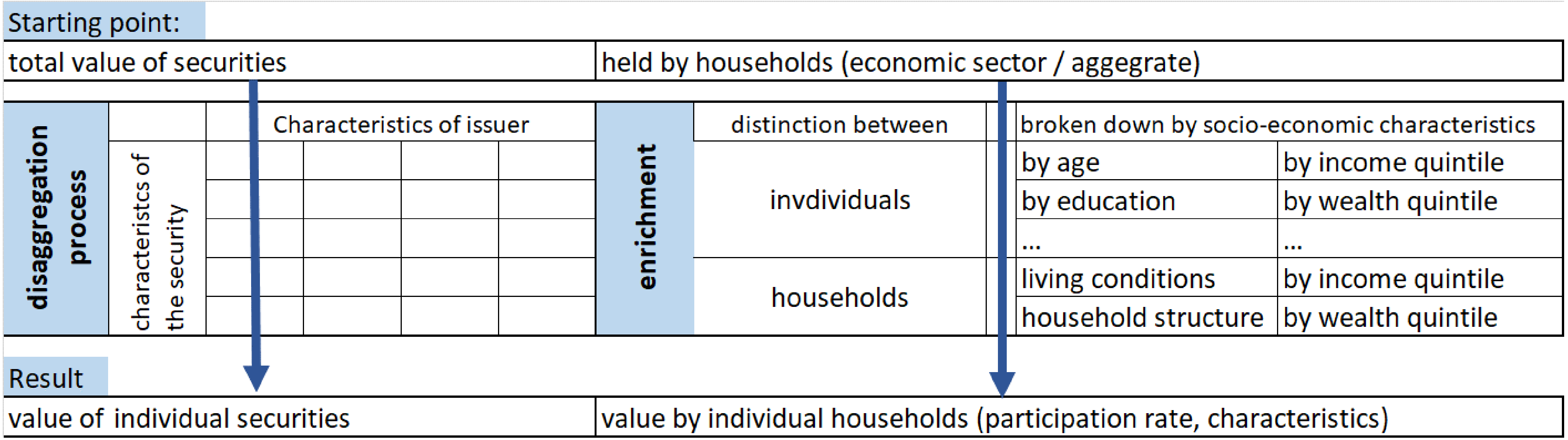
- Financial accounts data in Austria show shifts within deposits towards overnight-deposits and from debt securities to investment fund shares in most recent years against the background of nominal low interest rates.
- Financial accounts can easily explain which types of deposits or securities have been sold and bought
- but not: how many and which households are responsible for the portfolio changes?



Source: OeNB. *) Other equity, trade credits, non-life reserves, OTC financial derivatives and funds not yet paid

Drill down and enrichment shown in the case of securities

- Consequently, a link to distributional aspects in household survey data like HFCS can help to understand the financial behaviour and can be seen as enrichment of macro data by socio-economic characteristics in addition to disaggregate macro data as can be shown for securities



Agenda

- When analysing macro data using micro data as enrichment for socio-economic characteristics the question arise:
  - How comparable are individual (financial) instruments in HFCS compared with those of financial accounts (**methodological consistency**) ?
  - How comparable is the coverage of individuals (**scope consistency**) ?

- Based on our own experiences in the last ten years, the presentation will highlight the restrictions and benefits of such an enrichment in the case of Austria

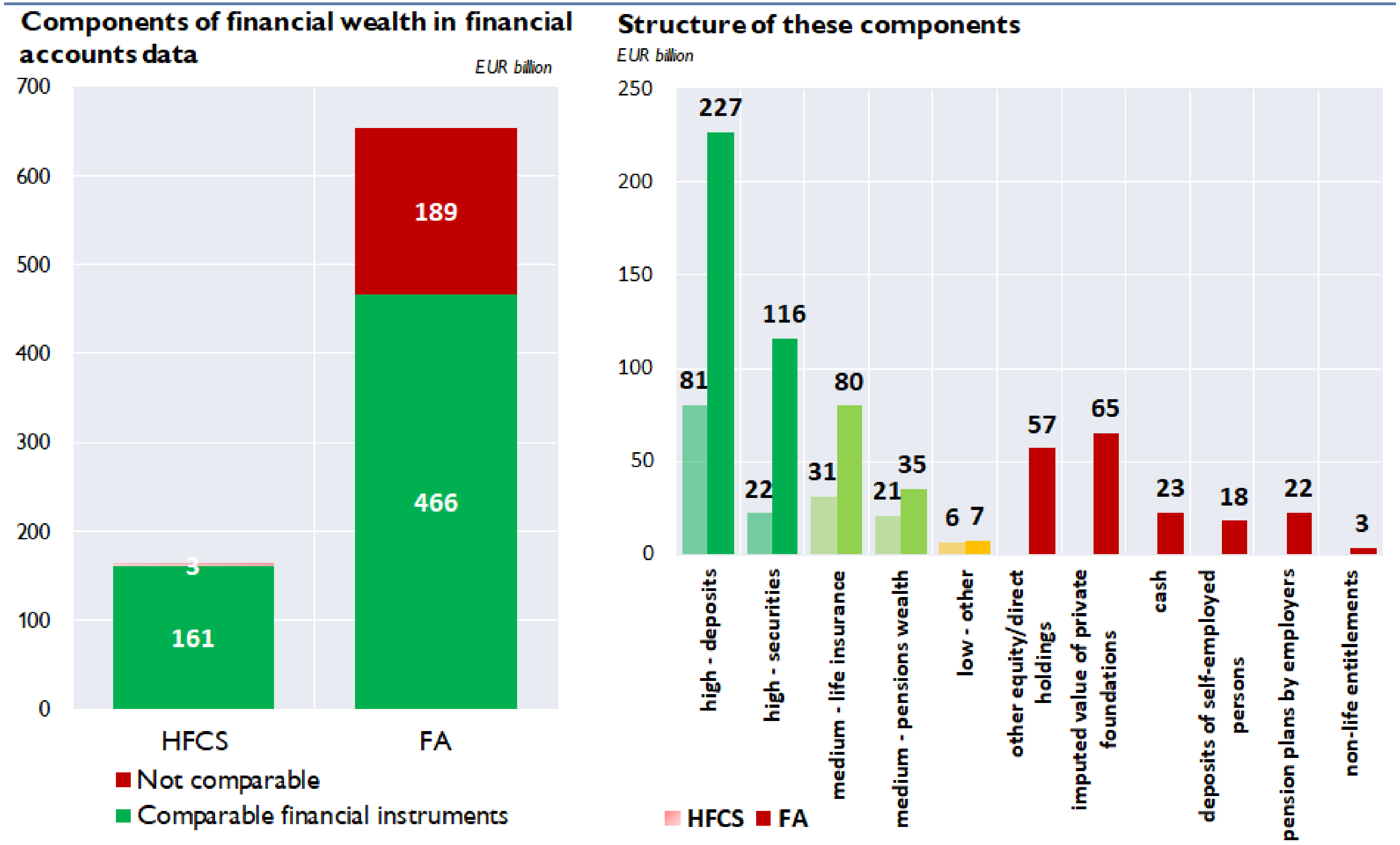
  - For details on HFCS in Austria see: Publication by the OeNB at www.hfcs.at; especially "First results for Austria" on the Eurosystem HFCS 2017, January 2019 (wave 3)
    For details on financial accounts in Austria see: "Manual on sources and methods for financial accounts data" published by the OeNB at www.oenb.at / statistics / data / financial accounts / downloads
  - For details on this exercise of comparison see: Micro and macro data: a comparison of the Household Finance and Consumption Survey with Financial accounts in Austria, Michael Andreasch and Peter Lindner; Journal of Official Statistics, Volume 32, Number 1, pp. 1-28, 2016

Conceptional comparability of financial instruments in HFCS and FA

Euro area and national quarterly financial accounts

NA/FA (ESA 2010)	HFCS	Conceptual comparability
FINANCIAL ASSETS (+)		
F21 Currency	N/A	N/A
F22+F29 Deposits	Deposits	High
F3 Debt Securities	Bonds and other debt securities	High
F4 Loans	Money owed to household	High
F5 Equity and investment fund shares	Shares, publicly traded	Medium to High
	Investment in non-self-employed business	
	Investment in self-employed business[7]	
	Mutual Funds	
F6 Insurance, pension and standardised guarantee schemes	Voluntary pension/whole life insurance schemes	Medium
	Occupational Pension Plans[8]	Low
F7 Financial derivatives and employee stock options	Other financial assets	Low
F8 Other accounts receivable		
N/A	Managed Accounts	Low
LIABILITIES (-)		
F4 Loans	Mortgages and loans	High
	Outstanding debts on credit cards, credit lines and overdraft balances	
F8 Other accounts payable	N/A	N/A
FINANCIAL NET WORTH		

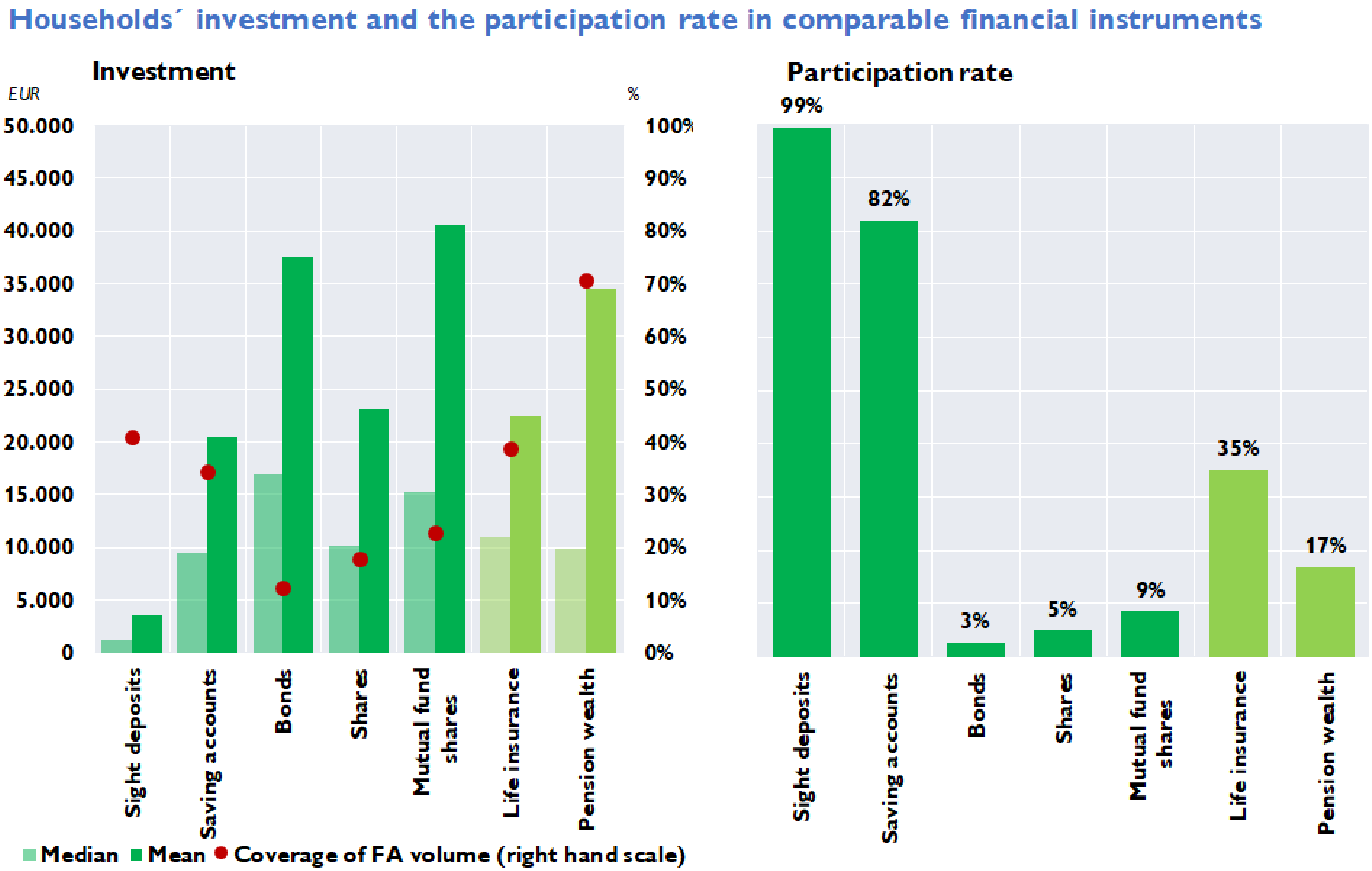
Financial accounts in Austria – value of comparable components



Quelle: OeNB.

- More than 70% of total financial assets held by households are comparable
- Within the comparable instruments, deposits dominate the picture (approx. 50%)
- Bonds, shares and mutual fund shares (securities) account for 25%
- Followed by medium comparable life insurance and pension entitlements

Structure of HFCS and FA data - Coverage of the volume in HFCS



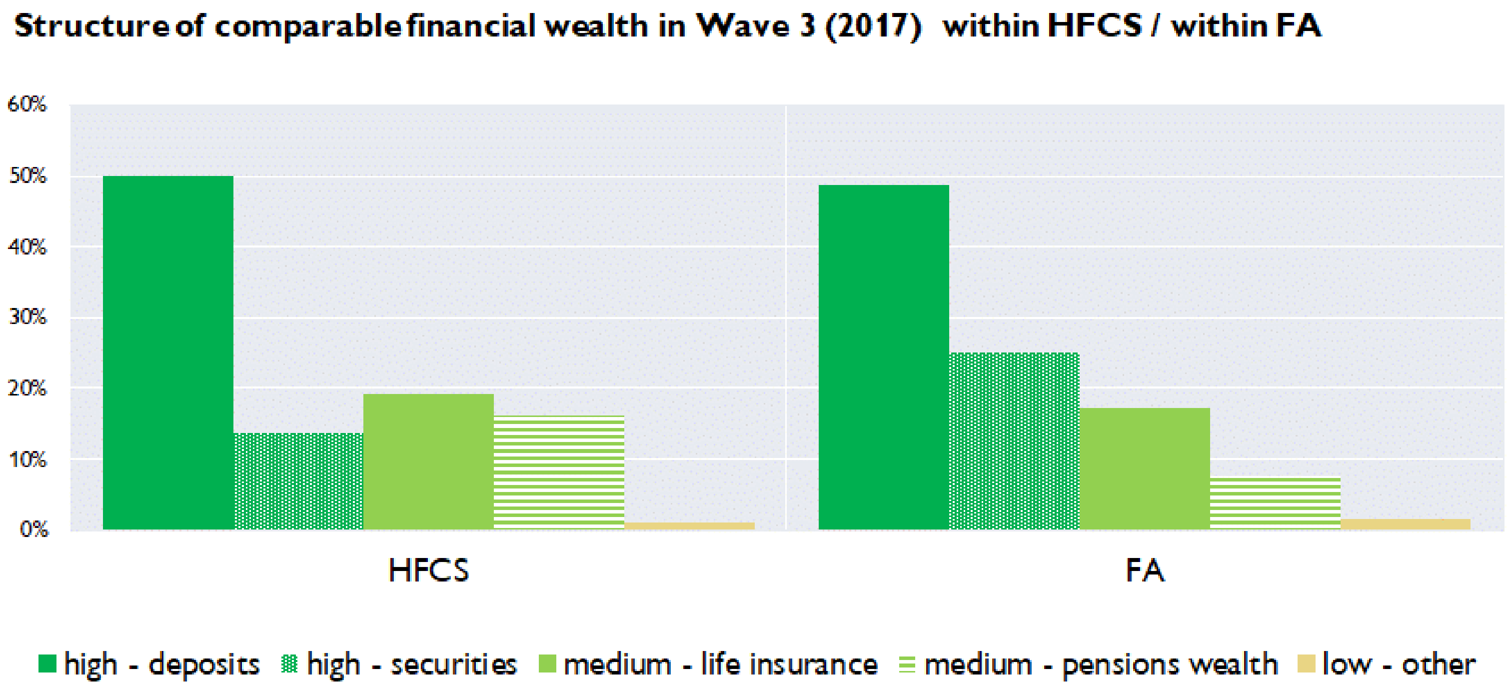
Source: OeNB (HFCS Austria, Financial Accounts)

- Obviously, nearly every household has an sight account with a bank with rather low investment
- Saving accounts are also widespread (participation rate 82%), the median value account for EUR 10,000 EUR.
- On the contrary, securities (bonds, shares mutual funds shares) are only held by less than 10% of households, but the mean value exceed the value for saving accounts
- Mean value of life insurance and pension wealth is close to the level of saving accounts, the participation rate stood at 35% and 17%, respectively

Structure of HFCS and FA data – Coverage of the volume in HFCS

ONB

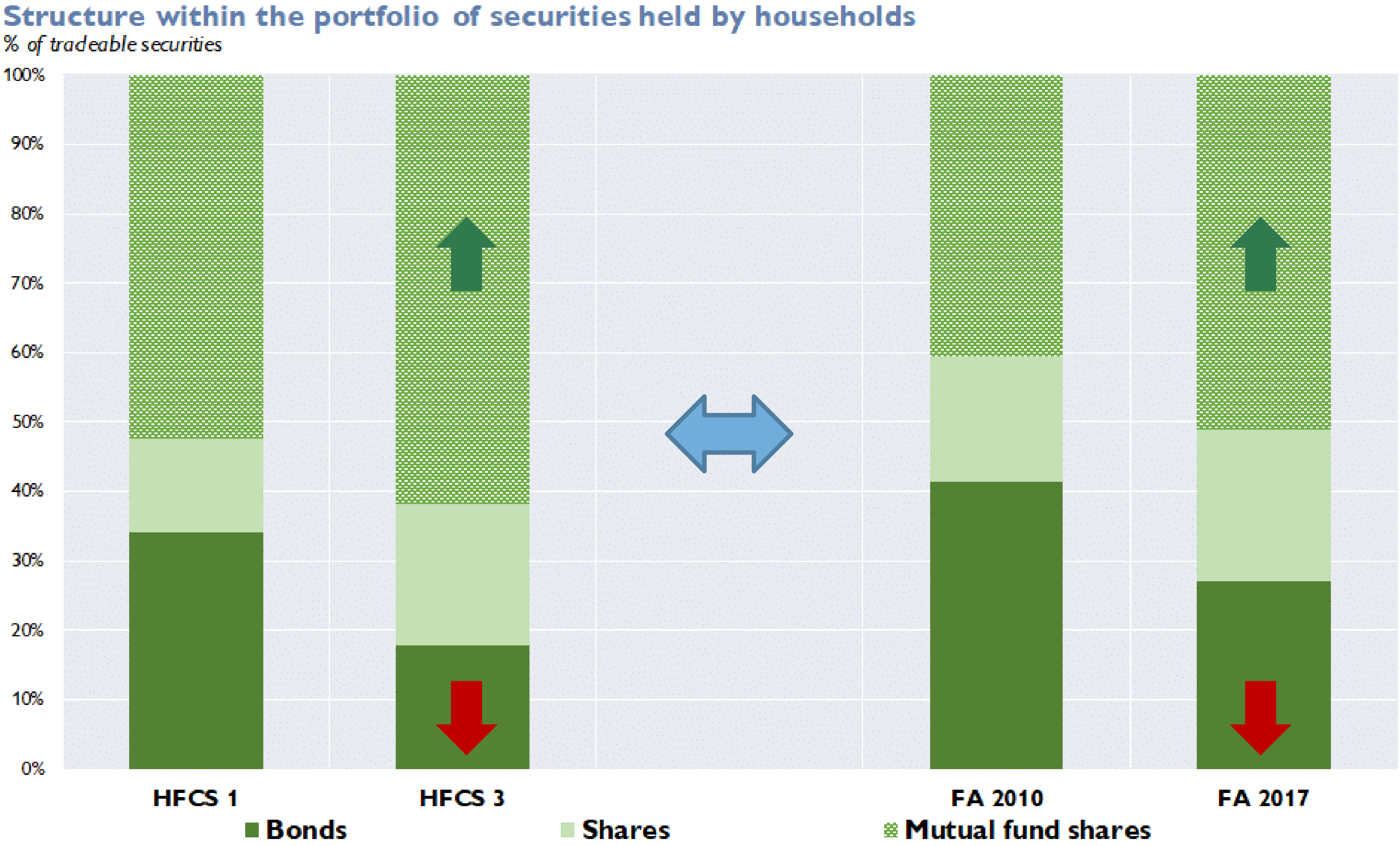
Composition of comparable financial wealth in HFCS and FA



Source: OeNB. HFCS Austria 1st wave 2010, 3rd wave 2017, FA = corresponding macro data in financial accounts

- Similar structure both in micro data and macro data allowing to bridge results from HFCS data derived from surveys every three years with financial accounts data (FA) compiled on a quarterly basis
- Deposits attribute more than 40% in both data sets
- Households held nearly at the same level (approx. 25%) securities and life insurance contracts.
- Together, highly comparable deposits and securities covered in HFCS data 64%, in the corresponding FA 75%

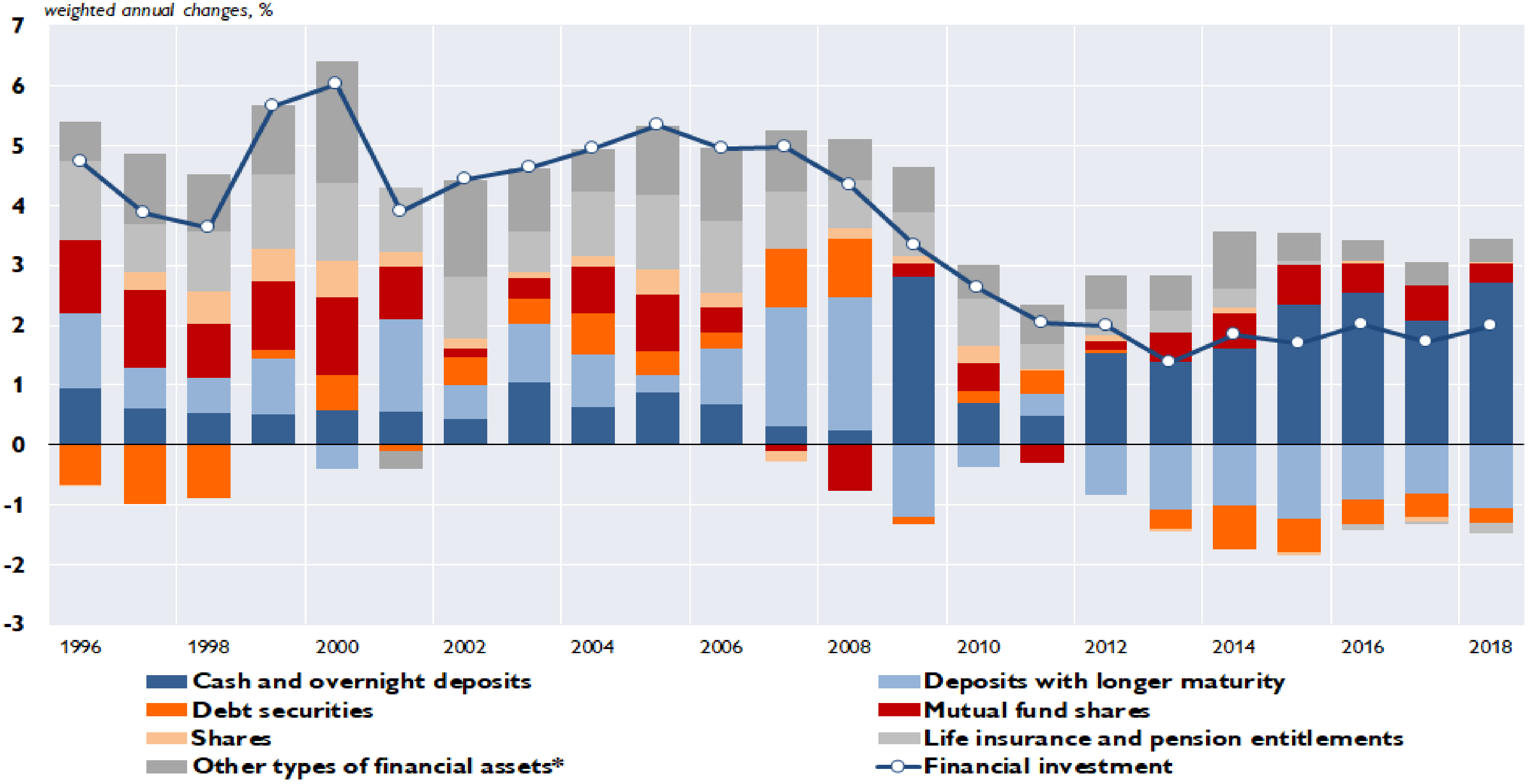
Bridging the information between HFCS and FA for securities



Source: OeNB. HFCS Austria 1st wave 2010, 3rd wave 2017, FA = corresponding macro data in financial accounts

- Both data sets (HFCS and financial accounts) show a significant decrease of the importance of bonds within the portfolio of securities (around 15 percentage points)
- Simultaneously, the proportion of mutual fund shares increased by around 10 percentage points.
- 80% of the securities are held by wealthy households (4th and 5th net wealth quintile)

Back to the recent development in the financial behaviour of households: Enhanced analysis using combined information



Source: OeNB. *) Other equity, trade credits, non-life reserves, OTC financial derivatives and funds not yet paid

- Indeed, due to comparability, we could identify that the shift from deposits with longer maturity to overnight deposits is a phenomena which can be observed for nearly all households (regardless of net wealth)
- While, less than 10% of the households are responsible for the shift from bonds to mutual funds shares (limited to households with higher net wealth)

Danke für Ihre Aufmerksamkeit

Thank you for your attention

www.oenb.at

oenb.info@oenb.at

@OeNB

OeNB

OeNB

Irving Fisher Committee on
Central Bank Statistics
BANK FOR INTERNATIONAL SETTLEMENTS

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Linking micro datasets to better service policy-making and analyses[1]

Jean-Marc Israel,

formerly European Central Bank

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Linking micro datasets to better service policy-making and analyses

Jean-Marc Israël
Ex Head of Analytical Credit & Master Data Division, ECB

The views expressed are those of the author

IFC-BNM-ECB Satellite (Session 2)

Kuala Lumpur, 16 August 2019

Content

1. ECB Statistics supporting a wide range of policy-making

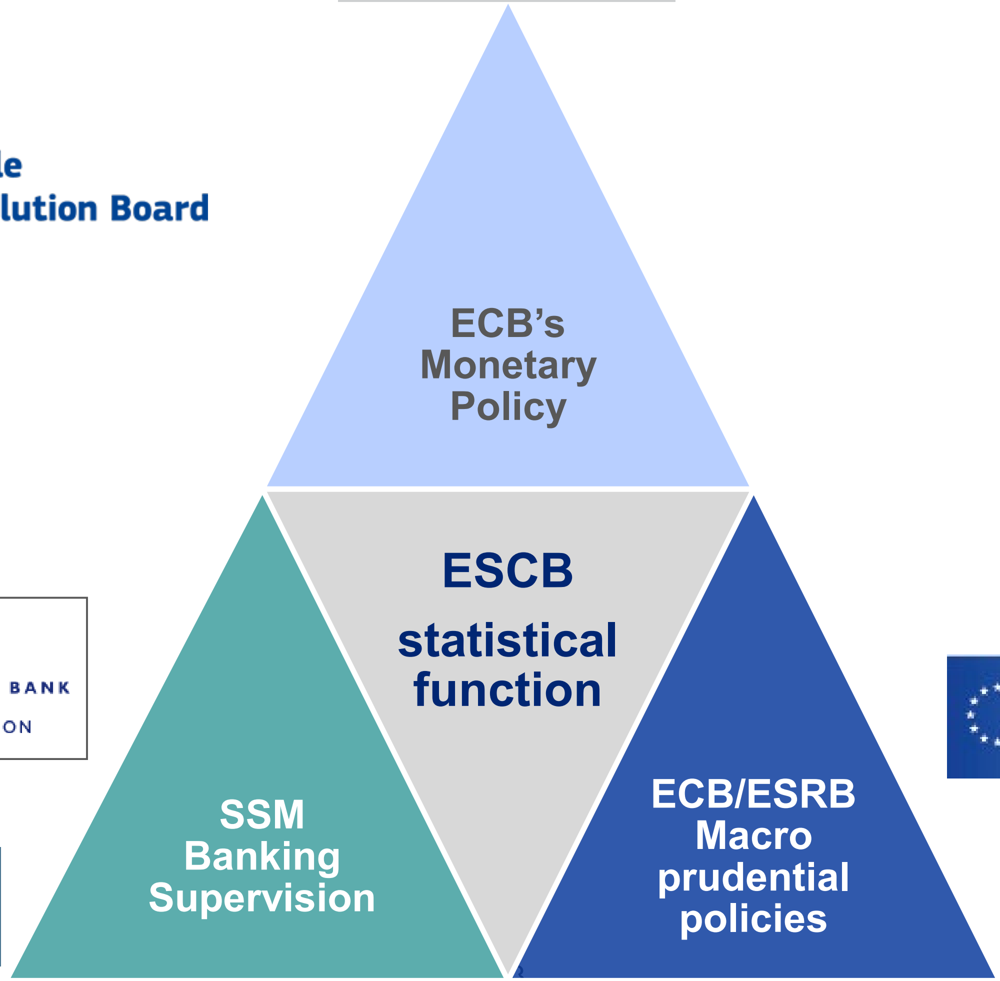
EUROPEAN CENTRAL BANK
EUROSYSTEM

Single Resolution Board

EUROPEAN CENTRAL BANK
BANKING SUPERVISION

EBA EUROPEAN BANKING AUTHORITY

ESRB



2. A magnifying glass for Central Banks' tasks (1/6)

"…well-established ESCB statistics will continue to provide the "big picture" of economic developments. But we should also offer a magnifying glass."

Mario Draghi, ECB President, 8th ECB Statistics Conference, 2016

✓ Moving *beyond and behind aggregated* data on *credit* to corporations and related *credit risk* using granular AnaCredit, *as a "magnifying glass"* to:

- Better understand **monetary policy transmission** and **systemic risk**
- Respond to **unforeseen policy needs**
  - ⚠ *Non-standard times => Non-standard measures => New statistics!*

2. A magnifying glass for Central Banks’ tasks (2/6)

Multipurpose dataset

Monetary policy conduct

Demand & supply in credit markets and access to finance of SMEs

- Evaluate lenders' risk exposures and balance sheet conditions and borrowers’ indebtedness to appropriately **identify *demand* and *supply* in credit market**
- Identify **shocks** *to demand* and credit crunches
- Analysis of *credit market conditions* **for SMEs**
- Assess the impact of non-standard monetary policy measures on SMEs access to funding through bank credit

Multipurpose dataset

Monetary policy implementation

- **transmission channels** of standard/non-standard monetary policy measures
- study "*credit channel*" and "*risk-taking channel*" of (standard) monetary policy
- assess non-standard measures such as liquidity provision (TLTROs) which are targeted to specific types of borrowers
- detailed information on *banks' loan portfolio* and evaluate the impact of measures on *firms' access to finance*, especially SMEs

2. A magnifying glass for Central Banks' tasks (4/6)

Multipurpose dataset

Risk management

Sectoral risk analysis/monitoring for macro stress testing

- detect *systemic vulnerabilities at sectoral level*, e.g. compute expected shortfalls, exposure indexes

- improve macro-stress testing, by allowing to develop a module for SME losses using micro data

Multipurpose dataset

Financial Stability

Risk exposures, interconnectedness and potential contagion

- study the *interconnection between host-country banks' loans* and their mother companies via the credit portfolio and credit risk associated

- understand *interconnectedness* and possible *contagion* across jurisdictions

- analyse *capital flows in periods of crisis* and assess possible policy instruments to mitigate liquidity dry-out

2. A magnifying glass for Central Banks' tasks (6/6)

Multipurpose dataset

Developing new and enhancing existing statistics

- Enhance the *quality of statistics* comparing aggregated and granular data; Compile new breakdowns of aggregated statistics not collected directly from reporting agents

- Further analyse credit allocation and *credit risk concentration* to assess their economic and employment impact

- Support and monitor new areas of policy, e.g. *green finance*

3. AnaCredit in a nutshell (1/6)

- AnaCredit = *Analytical Credit Datasets*
- *Loan-by-loan* information on (euro area) ***banks'*** *credit exposures* to *all legal entities* – including Small and Medium size Enterprises
- All euro area (19) and, nearly, all (27) *EU countries*
  - *Reporting* initially *from Credit institutions only*
- Basic features:
  - *88 data attributes per loan*, including lender and borrower identifiers (=>link to *business register*), credit and credit risk, interest rate
  - Reporting frequency: *monthly* (quarterly for some attributes)
  - *Proportionality* with possibility for NCBs to grant (full or partial) *derogations to smaller institutions* and € 25,000 *reporting threshold*
  - *First reporting* in November 2018 = 1st reference period *Sept. 2018*

3. RIAD in a nutshell (2/6)

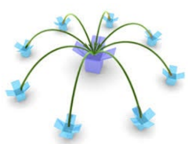
RIAD as the backbone of all ESCB granular statistics

- RIAD to serve as unique, *shared master data set serving all ESCB granular data collections*: lenders and borrowers (loans), issuer and holders (securities) and beyond
- A *shared platform*, for several stakeholders both in 'write' and in 'read mode', with NCBs as national hubs
- Advanced compounding rules to always derive the best information *at any point in time*
- Strict checks to ensure *high data quality and consistency*
- Flexible derivation of *group structures* based on different definitions

Comprehensive data model

incl. reference data on individual units plus relationships among them

A. Identification	e.g. *identifiers, address*
B. Stratification	e.g. *industrial activity, geographical allocation*
C. Demographic developments	e.g. *birth/closure date, corporate actions*
D. Relationships between units	e.g. *ownership, control,* (*fund*) *management*



Linking different datasets: **a key feature for data integration**!

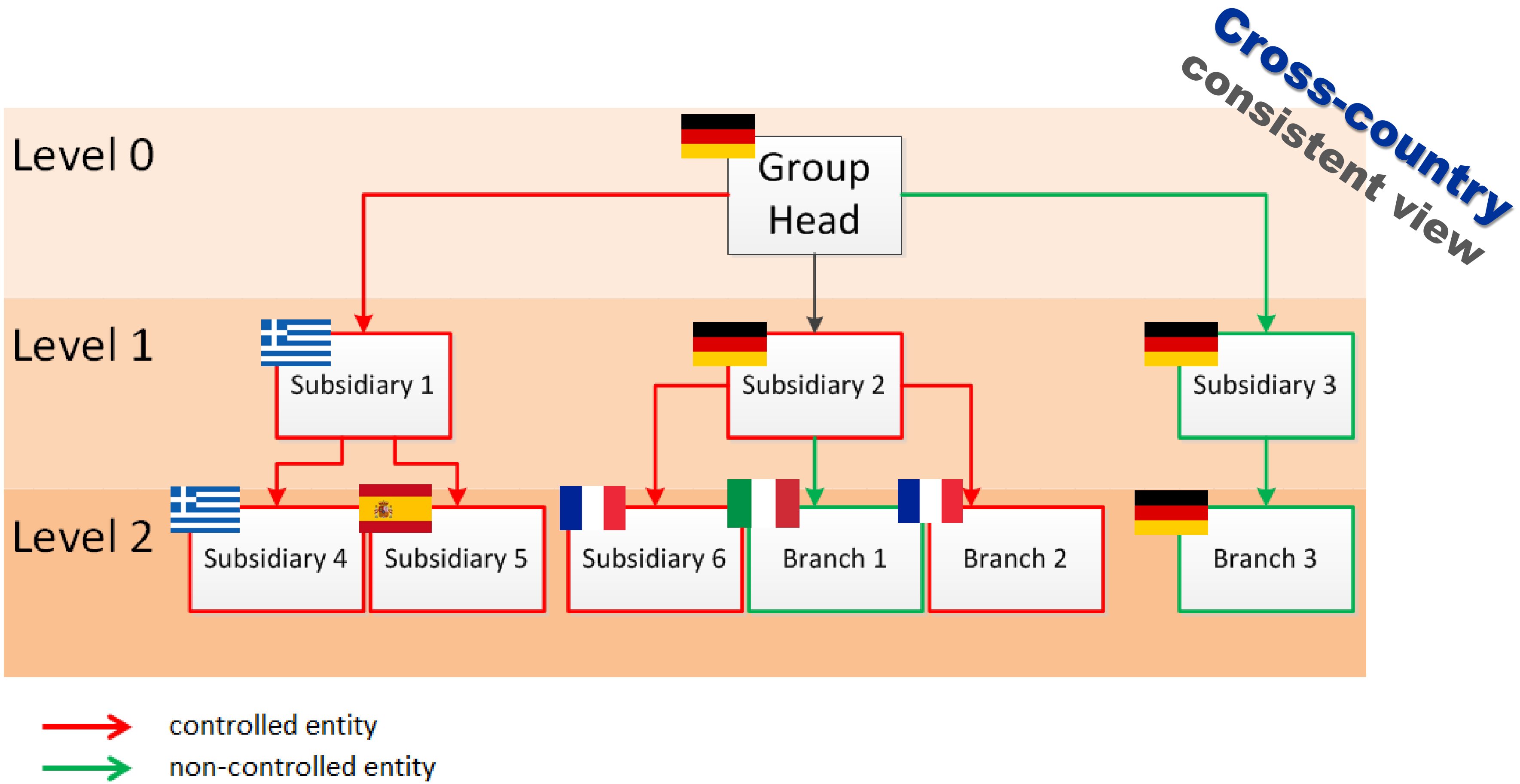
Full **historisation** of all data … ***new*** data versioning

Elaborate **access management**

e.g. *special application roles* and *different levels of confidentiality*

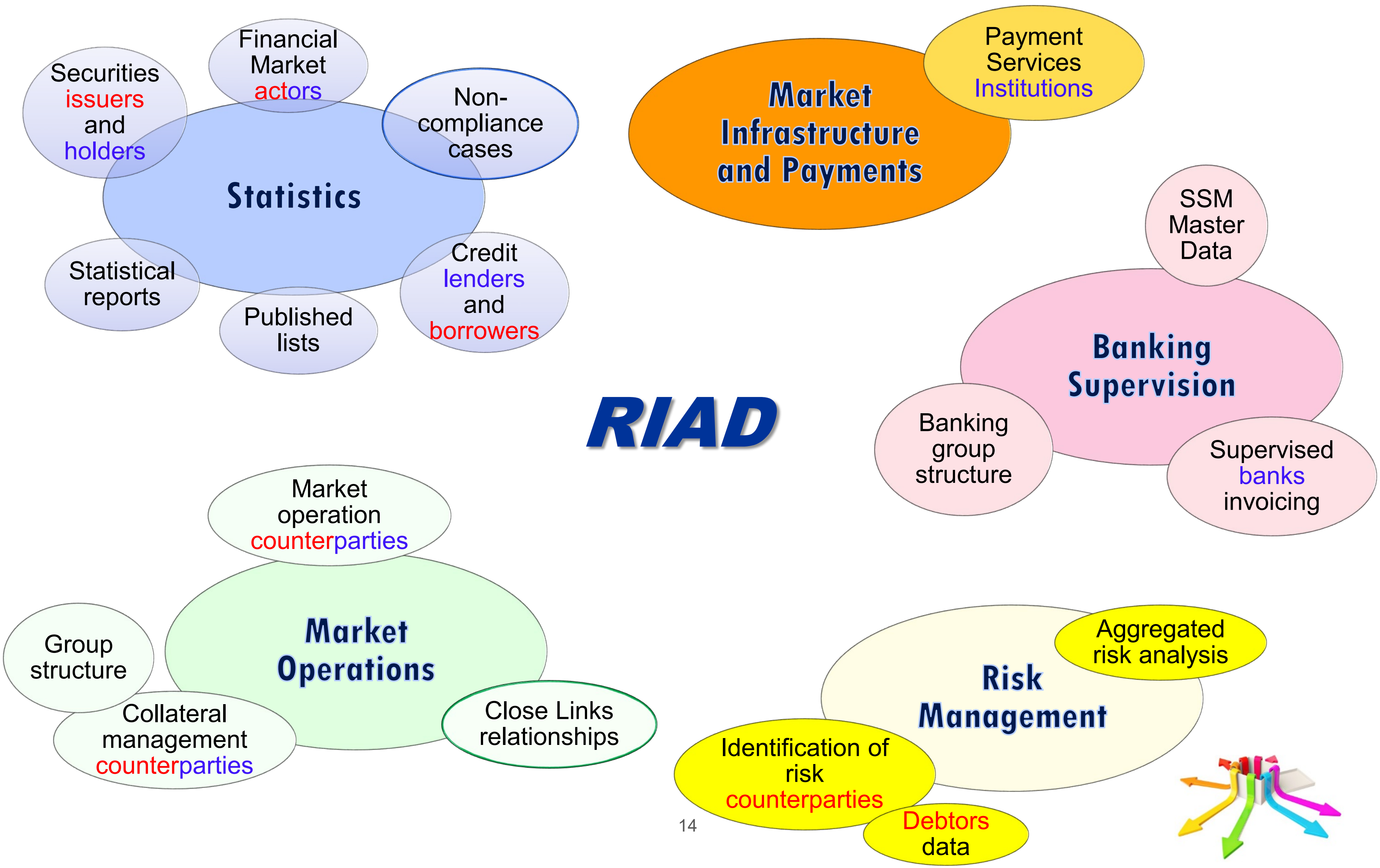
3. RIAD in a nutshell (4/6)

- RIAD allows to identify, store and present *two types of group composition*
  - ✓ Type A – based on direct and indirect "control"
  - ✓ Type B – based on "pure ownership" (all relationships)

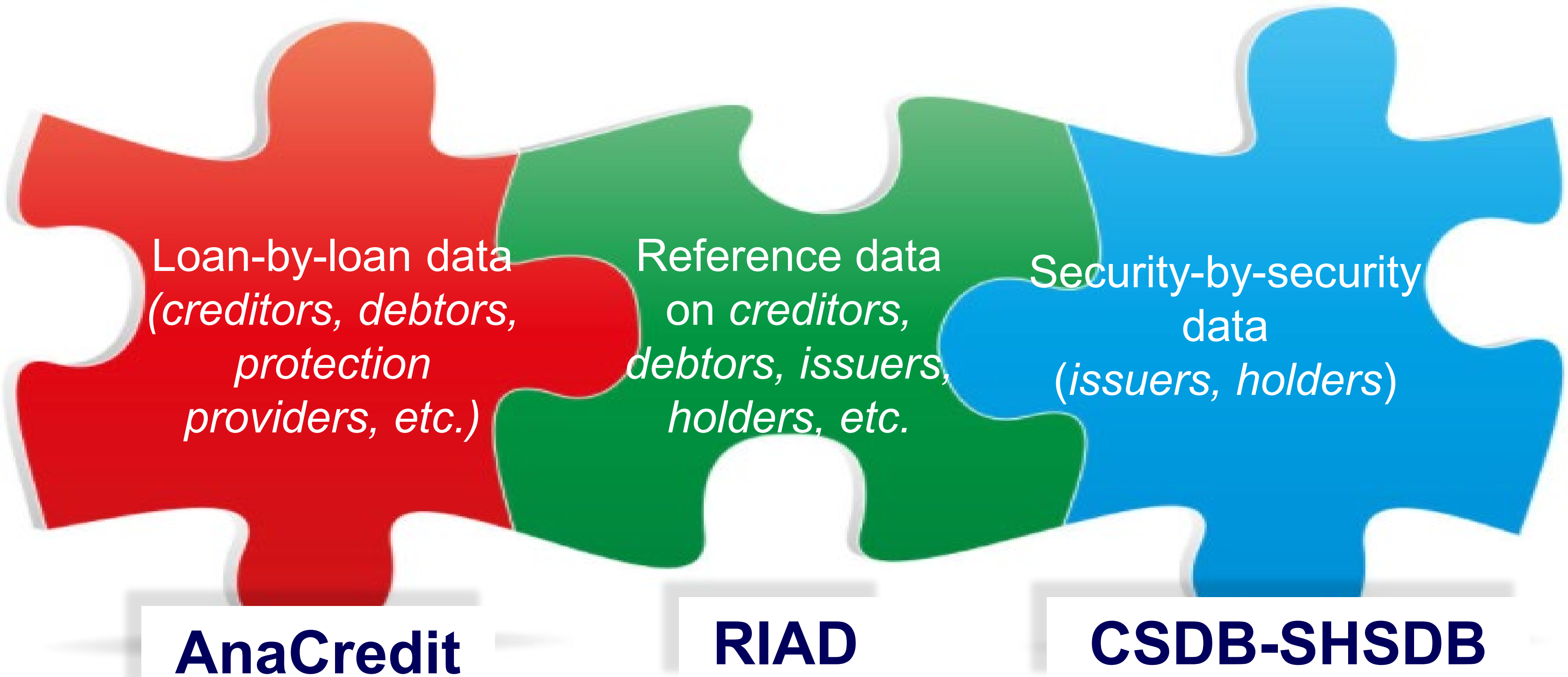


3. RIAD Stakeholders and Use cases (5/6)

Standardising and integrating existing frameworks *across domains:*



- To gradually increase availability of *granular data* at the ECB, while ensuring a *holistic approach* to data reporting building on
  - *Consistency:* same concepts and definitions across domains
  - *Interoperability*: possibility to combine granular information from different frameworks (e.g. securities, loans) to assess total exposures



4. Challenges and opportunities (1/3)

- […] ***big data*** also involve challenges related to, e.g., *identification* and *effectiveness and efficiency* in the data usage and analysis
- Dynamic factor models, or Bayesian shrinkage, can help address difficulties arising from the *multiple dimensions* of dataset
- But some methods are still being developed, e.g. for cases involving many observations – over time as our datasets grow ever larger

Speech by Benoît Cœuré, Member of the Executive Board of the ECB, at the conference on "Economic and Financial Regulation in the Era of Big Data" - Banque de France, Paris, 24 November 2017

The Eurosystem supports banks in implementing AnaCredit

AnaCredit Manual

- Detailed *documentation to clarify* the reporting for AnaCredit datasets
- To provide additional background and guidance with concrete cases/examples
- Parts I, II and III = 577 pages have been *published* a year-and-a-half prior to go live
- AnaCredit Manual complemented by Q&As
- Validation checks (also published) to help automate

Cooperation via BIRD

- Cooperation with the banking industry within ***Banks' Integrated Reporting Dictionary*** *(BIRD)*
- BIRD provides a formalised representation of AnaCredit as set out in Regulation *from the business point of view*
- Commercial banks participate on a voluntary basis

4. Challenges and opportunities (3/3)

- **Complex financial world** reflected in *complex AnaCredit data model* necessary to support a ***wide range of analyses on credit and credit risk****;* still the data model is *explained at length in the Manual and Q&As*

- Statistics areas will support researchers at central banks (and beyond) by offering *pre-defined 'views' of the data* tailored on their needs

- Continuous **dialogue** *with researchers and other users* to define most appropriate *data marts* for pre-defined queries and analysis; Also **banks** will benefit from *feedback loops*

- Ensure **methodological support** – e.g. consolidate exposures or debt –, *correct usage of the data* and interpretation of the results

5. The way forward

- First delivery of data to users – expected *by end-2019*
  after ca. 12 months to fix initial teething problems

- Acquire and further develop tools to *analyse the data*
  with a view to deriving stories out of them

- Then, start reflecting on possible extensions – *2022 and beyond*
  - *Extend coverage to other lenders, e.g. FVCs, FCLs, Insurance*
  - *Extend credit risk attributes*
  - *Cover other instruments*

- Ensure more *data integration* with other reporting
  e.g. banks' balance sheet (BSI), banks' interest rates (MIR),
  securities holdings (SHS) from all sectors or large institutions

Thank you!

Questions?

Word-cloud from the AnaCredit Regulation

ECB broader strategy for statistics: key words

Harmonisation based on Standards

define common practices and processes for data production based on standardised concepts

Integration

managing various areas of statistical (and supervision) information as parts of a single system

Digitalisation

benefit from dematerialisation of financial documents in the banks' systems

2. A magnifying glass for Central Banks' tasks

- AnaCredit will provide *complete* and *harmonised* information on credit and credit risk for all euro area countries - and beyond…
- Data based on
  - Common *concepts and definitions* across 19, and actually 27, countries
  - Unique *reporting threshold* - € 25,000
- Comprehensive and harmonised *data assurance metrics*
- Gradually *replacing existing credit registers*
- A *10-year-old dream* comes **true**!

"For a more resilient international financial system we need a global credit register based on a harmonized approach with adequate standardization across countries" *Issing Committee* (*February 2009*)

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Quality checks on granular banking data: an experimental approach based on machine learning[1]

Fabio Zambuto,

Bank of Italy

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

BANCA D'ITALIA

Quality checks on granular banking data: an experimental approach based on machine learning

Fabio Zambuto

Bank of Italy

Statistical Data Collection and Processing Directorate

16th August 2019

Outline

- Context and Motivations
- Data
- The Algorithm
- Results
- Conclusions and Future Work

Context and Motivations (1)

- Central Banks collect, process and disseminate a wide set of statistical data: Data Quality Management (DQM) is crucial to support decision making.

- DQM in Bank of Italy: **automated** checks to verify **pre-determined relationships** in the data (e.g. accounting, logical and mathematical relationships).

- When deterministic relationships are **weak** DQM entails plausibility checks (trend-based) that rely on “acceptance regions” to isolate outliers.

Context and Motivations (2)

- Shortcomings of plausibility checks:
  - Calibration not straightforward
  - Periodical revision and update needed
  - Large number of acceptance thresholds.

- Complex and time-consuming system with **highly granular** data and **heterogeneous** reporting patterns.

- **Aim**: explore the use of ML techniques to improve plausibility checks in granular databases.

- Approach: a supervised learning algorithm (Quantile Regression Forests) employed to detect potential outliers.

Findings

- Application to payment services data reported by banks. Outliers cross-checked with reporting agents.

- Empirical results:
  - **New outliers** detected (not identified by the current DQM system).
  - **High accuracy** (77% precision; reduced “false positives”).

- Improvements:
  - Thresholds tailored to the characteristics of banks and to the degree of granularity of the data.
  - Dynamic thresholds that are automatically updated as new data are reported. Reduced involvement of analysts.

Data

- Focus on debit cards issued:
  - Unit of analysis = n. of cards issued by bank (***i***), at the end of the semester (***t***), for a given province (***p***).
  - Data extracted from DWH. Period: Dec-2014 to Jun-2018.

- Additional data on bank features:
  - n. of customers by province of the counterparty,
  - type of customer accounts,
  - other payment services offered (business model).

- Final sample: **18,000 observations** corresponding to **213** banks.

The Algorithm (1)

- Analysis of the empirical distribution of the n. of debit cards (Y) conditional on bank characteristics (Xs).

- Estimation of quantile functions $q_\tau\ (Y|X)$:

$$Prob(Y < q_\tau(X)) = F(q_\tau(X)) = \tau$$

- Quantile functions combined to form **prediction intervals** (acceptance thresholds) associated with a given probability (α):

$$PI(X) = [q_{\frac{\alpha}{2}}(X), q_{1-\frac{\alpha}{2}}(X)]$$

- <u>Outliers</u>**:** values outside the intervals; unlikely to occur (too high/too low) given the reporting context.

The Algorithm (2)

- Sampling:
  - **Train** set to estimate **quantile functions** $q_\tau(x)$ for different τs.
  - **Test** set to compute **intervals** $[\hat{q}_{\tau 1}(x), \hat{q}_{\tau 2}(x)]$ and detect **outliers**.
- Training:
  - Algorithms: Quantile Regression Forest, Linear Quantile Model, Linear Quantile Model with Fixed-Effects.
  - Model selection with 10-folds cross validation.
- Testing:
  - Rolling window with two snapshots of data. Last two semesters in each snapshot as test set.
  - Outliers communicated to banks for cross-check.

The Algorithm (3)

- Model:

$$q_\tau(x_{ipt}) = \beta_0 + \beta_1 depositors_{ipt} + \beta_2 perc\_ca_{ipt} + \beta_3 size_{it} + \beta_4 iss\_acq\_ratio_{it} + \beta_5 trend + \beta_6 sem + \alpha_i + \mu_p$$

- Predictors:
  - $depositors_{ipt}$= N. of depositors (of a bank in a given province)
  - $perc\_ca_{ipt}$= % of depositors with current accounts
  - $size_{it}$= Total transacted amounts (as an issuer and as an acquirer)
  - $iss\_acq\_ratio_{it}$= Balance between issuing and acquiring services
  - sem = Semester dummy
  - $trend$ = N. of semesters starting from the first period in the dataset
  - α_i= Bank fixed effects
  - μ_p = Province fixed effects

The Algorithm (4)

- Estimated acceptance thresholds:

$$PI_1(x) = [q_{0.01}(x), q_{0.99}(x)]$$

$$PI_2(x) = [q_{0.025}(x), q_{0.975}(x)]$$

$$PI_3(x) = [q_{0.25}(x) - 1.5 \cdot (q_{0.75}(x) - q_{0.25}(x)), q_{0.75}(x) + 1.5 \cdot (q_{0.75}(x) - q_{0.25}(x))]$$

- Observations falling outside **any** of the intervals flagged as potential outliers.

Results

Cross check of outliers with banks

Prediction intervals:	PI_1 $[q_{0.01}, q_{0.99}]$	PI_2 $[q_{0.025}, q_{0.975}]$	PI_3 Inter-quartile range
a-Total number of potential outliers	**373**	**489**	**457**
b-Anomalies detected and revised ("true positives")	289	312	292
c-Confirmed observations ("false positives")	84	177	165
d-Precision b/a (%)	**77.5%**	**63.8%**	**63.9%**

Concluding Remarks

- Potential to improve DQM: more precise quality checks to detect outliers at a fine grained level with reasonable level of accuracy.

- Maintanance of DQM system: dynamic thresholds and periodical training of the algorithm vs manual update of acceptance thresholds.

- Additional challanges:
  - New processes and IT solutions for the production phase.
  - Communication of anomalies to banks becomes more complex.

Future Work

- Extensions:
  - Application to other payment services data (e.g. credit cards).
  - Analysis of data at the collection stage (i.e. before delivery to the DWH).
  - Classification algorithms (exploiting variations to reported data).
  - Unsupervised algorithms for outlier detection.
- In perspective: extend the ML approach to other granular data collections (in particular when current checks are weak).

Thank you for your attention!

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Predicting economic growth using machine learning techniques and sentiment analysis[1]

Berkay Akisoglu,

Central Bank of the Republic of Turkey

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

TÜRKİYE CUMHURİYET MERKEZ BANKASI

IFC- BNM- ECB

Predicting Economic Growth Using Machine Learning Techniques and Sentiment Analysis

Berkay Akışoğlu
Statistics Department , Central Bank of Turkey

16 August 2019 | Malaysia

Outline

I. Motivation

II. Methodology

III. Data

IV. Machine Learning Models

V. Sentiment Index

VI. Future Works

I. Motivation

- Time lag in macro-economic statistics
- The problems with regular surveys
- New data sources to monitor economic developments
- Financial and economic news

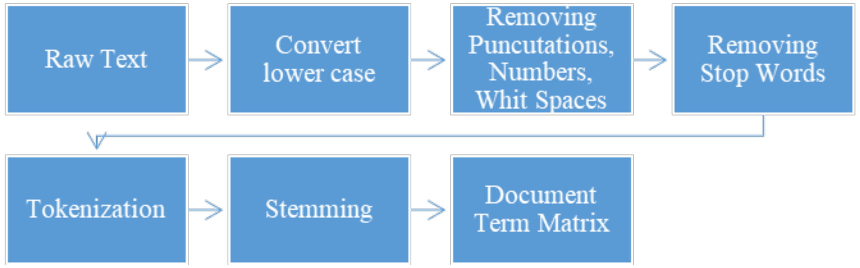
II. Methodology

- Sentiment Analysis: A field of research aiming to determine subjective information such as emotion, opinion, attitude stated in the text by using methods and techniques from fields such as natural language processing (NLP), statistics and computer science.

- Methods used in sentiment analysis
  - Lexical Approach
  - Machine Learning Models

II. Methodology

- Preprocessing Steps:
  - Convert lower case
  - Removing punctuations, number and white spaces
  - Removing stop words
  - Tokenization
  - Stemming



- Document Term Matrix

Doc 1: «**TÜRKİYE genelinde konut metrekare fiyatları nisan sonu itibarıyla son bir yılda…..»**

Vector-space representation

Docs	artış	aya	Fiyat	konut	liraya	metrekare	nisan	son
1	2	1	2	3	2	3	2	3

III. Data

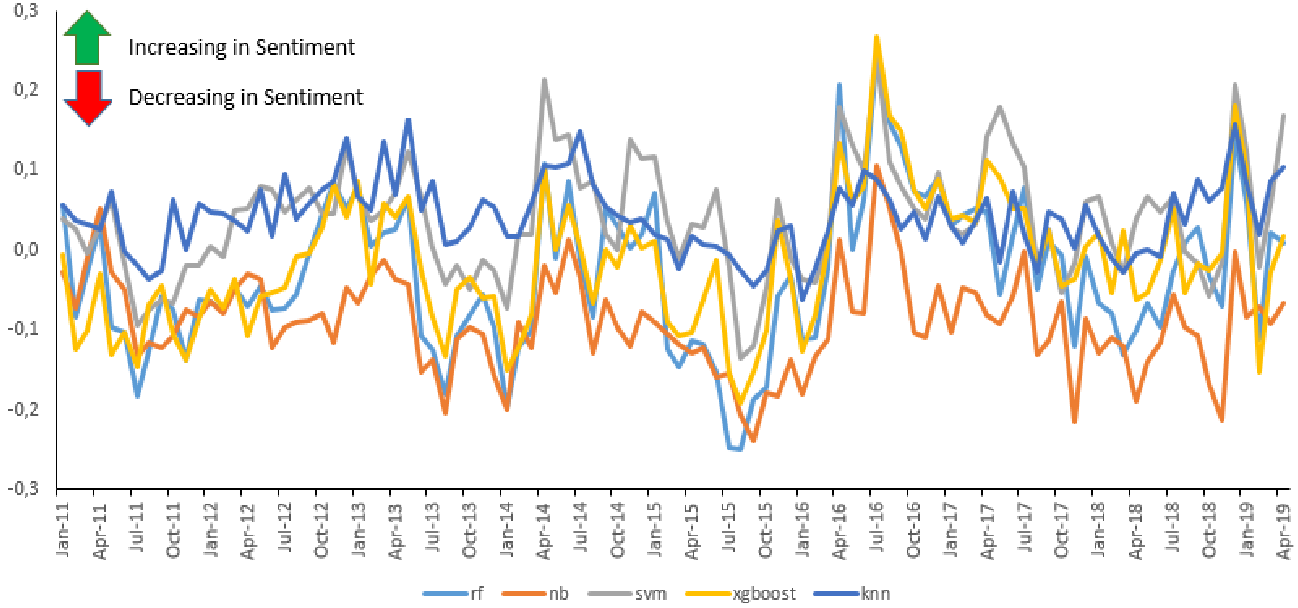
- Total 131601 financial and economic news
- Each includes name and phrases specified by Central Bank
- Mainstream and local media

Year	Number of News (as of April 2019)
2011	18.474
2012	21.042
2013	14.978
2014	15.400
2015	16.606
2016	15.227
2017	13.736
2018	13.397
2019	2.741
Total	**131.601**

III. Data

- No ready-to-use dataset whose sentiment orientation is known.
- Randomly generated news bundles were evaluated by Central Bank employees, academicians, graduates and students of TED University.
- Question: «What is your expectations of future economic activity after reading this news.»
- 1357 news labelled (three labels: «-1», «0», «1»)

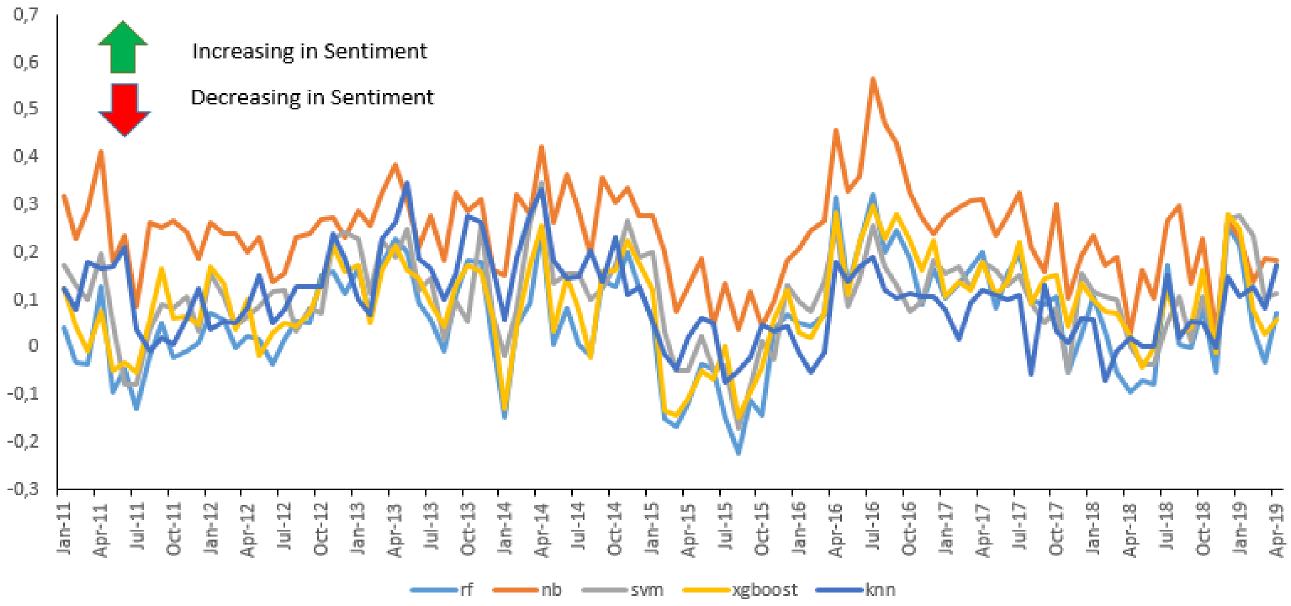
IV. Machine Learning Models and Sentiment Indexes (3 labels; -1,0,1)



Algorithm	Accuracy (%)
Random Forest	0.51
SVM	0.44
Naive-Bayes	0.47
KNN	0.43
XGBoost	0.45

	rf	nb	svm	xgboost	knn
rf	1.00	0.72	0.70	0.87	0.60
nb	0.72	1.00	0.60	0.55	0.50
svm	0.70	0.60	1.00	0.76	0.59
xgboost	0.87	0.55	0.76	1.00	0.51
knn	0.60	0.50	0.59	0.51	1.00

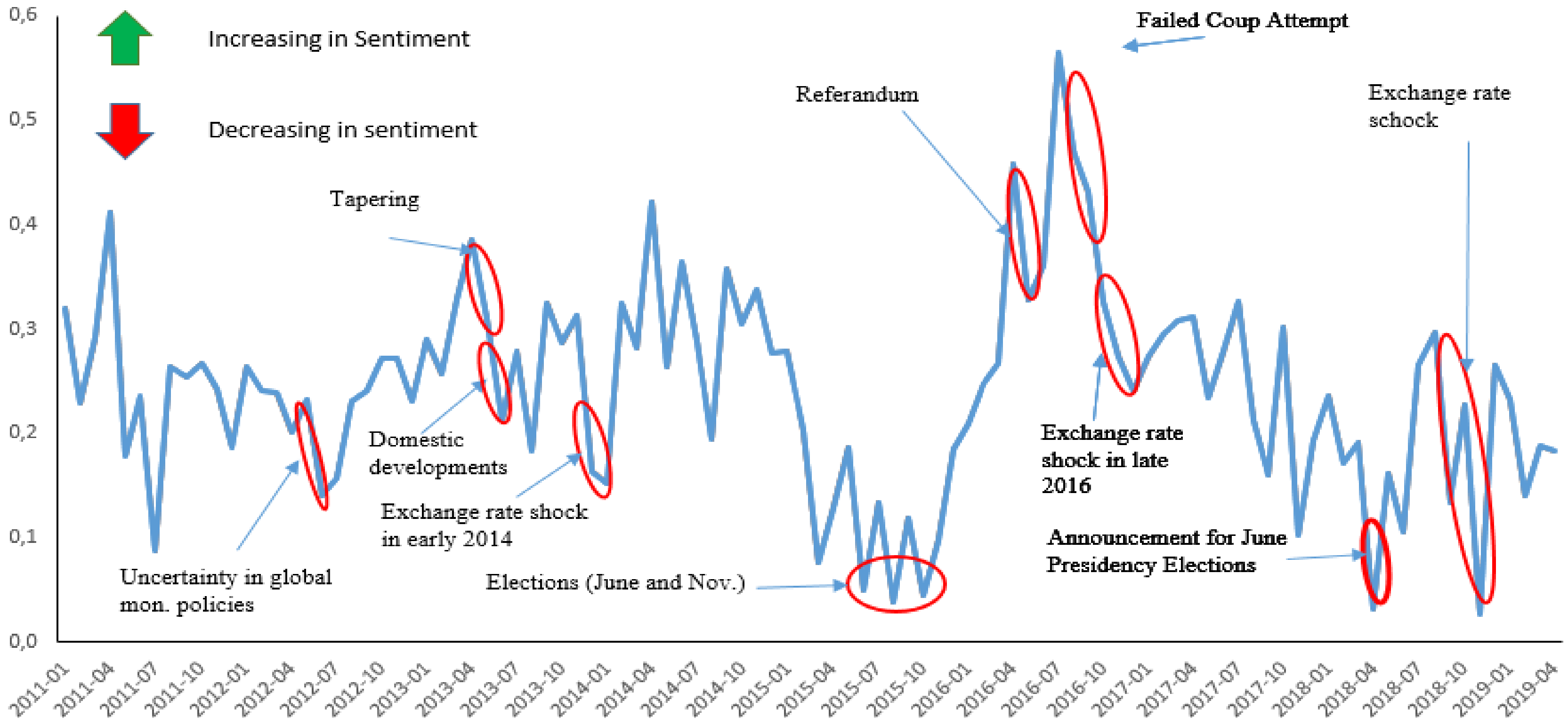
IV. Machine Learning Models and Sentiment Indexes (2 labels;-1,1)



	Accuracy	Precision	Recall
Random Forest	0.65	0.65	0.67
SVM	0.61	0.62	0.58
Naive-Bayes	0.61	0.59	0.66
KNN	0.57	0.55	0.66
XGBoost	0.61	0.60	0.66

	rf	nb	svm	xgboost	knn
rf	1.00	0.80	0.78	0.98	0.57
nb	0.80	1.00	0.69	0.73	0.54
svm	0.78	0.69	1.00	0.78	0.57
xgboost	0.93	0.73	0.78	1.00	0.49
knn	0.57	0.54	0.57	0.49	1.00

VI. Interpretation of Sentiment Index for Turkish Economy



Conclusion and Future Works

Conclusion

- First study using this techniques on Turkish financial and economic news
- Captures important economic and political events

Future Works

- Improving stemming
- Differantations between news
- Conducting survey with larger participation
- Running different algorithms on the same data set
- Trying econometric models to get most benefit from sentiment index
- Investigate the relations with other economic indicators

THANK YOU

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Granular data and stress testing: stepping up to the challenge[1]

Cees Ullersma and Iman van Lelyveld,

Netherlands Bank

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Forthcoming in:

Handbook of Financial Stress Testing, editors Doyne Farmer, Til Schuermann, Alissa Kleinnijenhuis, Thom Wetzer. Cambridge University Press

Granular data offer new opportunities for stress testing [1]

Cees Ullersma and Iman van Lelyveld

Main messages

- Data is becoming increasingly more available: more granular, more timely and reported more efficiently. In combination with high quality metadata, this trend allows for connecting and viewing the datasets in new and more informative ways.
- Such developments are changing how policymakers and regulators can operationalise stress testing. It allows stress testing to move towards modelling the complete financial ecosystem. It also enables (macroprudential) stress testing with micro-foundations, taking into account the diversity of market participants. The impact of shocks can be analysed from a highly aggregated macro level all the way down to individual market transactions and participants.

Abstract

Nowadays, more and more granular data is being collected. The 2008-2010 crisis has shown that authorities were missing crucial information for accurately identifying risks in the financial system. This realization has led to a significant increase in the depth and scope of information being reported across the system. At the same time, the cost of reporting has gone down due to further digitalization.

In this paper we will examine how increased granularity allows authorities to put the data to better use and we will focus in particular on stress testing. Stress testing involves postulating severe scenario's that often have yet to materialise. More granular data combined flexibly from different sources allow for more accurate assessment of such adverse hypothetical states of the world. For instance, detailed information on interbank exposures allows for understanding bank to bank contagion much better.

We will briefly discuss the developments in the last decade and discuss what possibilities arise. The new granular data sets allow us to implement stress tests on various levels of aggregations using the very same data sets. Then we will turn to prerequisites in terms of for instance meta data allowing to connect data sets and data governance specifying how to treat the data. Finally, we will discuss remaining challenges, ranging from data quality up to high performance IT infrastructure and a company culture that embraces data.

[1] We are grateful to Roos van den Berg and Jiaqi Zheng for their assistance in constructing the figures and to three anonymous referees whose comments have improved the chapter significantly.

Introduction

Nowadays, more and more granular data is being collected. The 2008-2010 crisis has shown that authorities were missing crucial information for accurately identifying risks in the financial system. For example, due to the over-the-counter (OTC) nature of derivatives markets there is no centralised overview of the market.[2] Participant only observe their own volumes and exposure concentrations. The major US investment banks therefore did not realise that *jointly* they were massively exposed to a single entity, the lightly regulated insurer AIG. In setting their capital buffers and implementing other risk mitigating procedures they were thus ignoring an important yet unobserved concentration risk.

The realization that more information is needed to properly capture the risk in the financial system, has led to a significant increase in the depth and scope of information being reported across the system.[3] At the same time, the cost of reporting has gone down because more and more economic interactions are recorded digitally. Also, as a side effect of internet based activity (i.e., commercial and social networks) new aspects of economic activity have become measurable.[4] Furthermore, the cost of analysis has gone down as well as (open source) tools and methods (e.g., machine learning (ML) and artificial intelligence (AI)) are improving markedly. Not only is more data available for stress testing than ever before, implementing more accurate models is within reach.

The spectacular growth in the granularity of data collected in the last century can be illustrated by comparing the level of detail available in Foster (1922) with that in Ehlers (2018), shown in Figures 1 and 2, respectively. Arguably Foster did not intend to depict each individual transaction contained in the flows but the data was also simply not there. For the flows shown in Figure 2 the disaggregated information is in principle available. Combining this information is another matter which we will discuss in detail below.

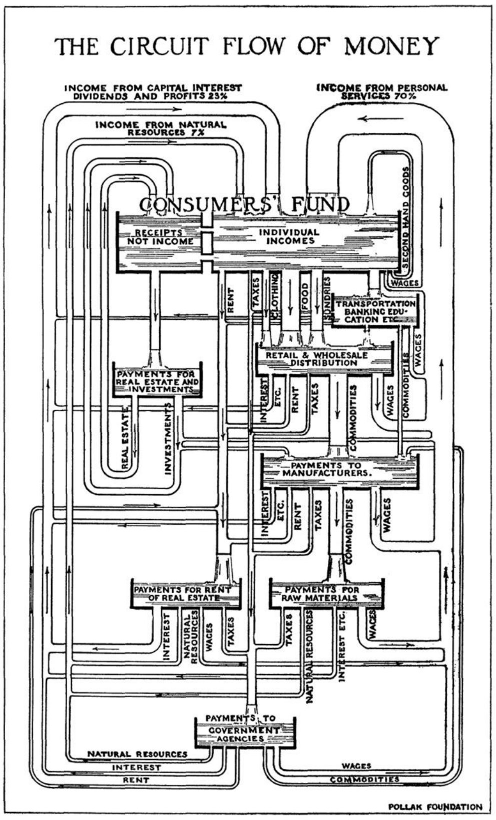
---

[2] See Abad *et al* (2016) for a first effort to shed light on these 'dark markets' and Levels *et al* (2018) for further detail.

[3] See for example the second progress report on the implementation of Phase 2 of the G20 Data Gaps Initiative (FSB and IMF (2017)).

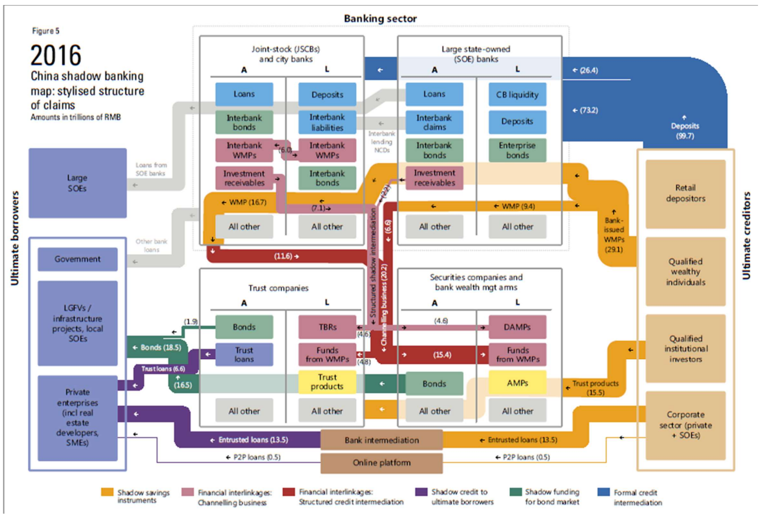
[4] See for example Graff (2018) who shows Brexit related worker migration in and out of the UK in a comprehensive and timely way using LinkedIn profiles.

Figure 1 Granularity 100 years ago: the circuit flow of money



Source: Foster (1922)

Figure 2 Granularity today: the Chinese financial system



Source: Ehlers *et al* (2018)

Data collection and analysis has been a core competency for most central banks and supervisors for many years. Central banks thus have a deep understanding of how to collect data from – primarily financial – economic agents and are thus well placed to capitalise on the current data trend. At the same time, they are faced with legacy systems that might lock them in inefficient solutions. Some authorities are looking for alternative ways to acquire data. In Austria, for example, direct reporting to the central bank has been replaced by reporting to a joint venture – owned by the bigger banks – tasked with data handling. This joint venture, in turn, provides the necessary reporting to the authorities (Kienecker (2018)). An alternative is plugging into the systems of private institutions directly. In theory, this would allow authorities to extract the precise data points at precisely the right time without imposing material reporting burden. The practice is more elusive because firms have multiple IT systems with unaligned definitions. For regulatory reporting, significant manual effort is required. Direct access to firms' data will not solve this but will push the necessary manual steps to the authorities.

The breadth of reporting has increased especially in the area of Over-the-Counter (OTC) markets. In OTC markets, such as the interbank market or the derivatives markets, market conditions are more difficult to gauge since contracts are bilateral and generally private information. Recently regulators have introduced new regulation to – at the minimum – introduce exposure reporting (without imposing limits or other constraints). The European Market Infrastructure Regulation (EMIR) is an example of such a regulation covering derivative markets trading. Under this regulation all parties with significant dealings need to report their transactions centrally.[5] These central repositories are for-profit entities that collect the data and disseminate it to the relevant authorities. Another important initiative is AnaCredit which has become operational in the fall of 2018 and will provide a basic version of a European credit register. This will allow authorities to more comprehensively assess credit risk. Furthermore, AnaCredit data can be used to trace out the effects of monetary policy. A final example is the Money Market Statistical Reporting (MMSR) covering large banks and their counterparties in the European money market. It contains deal level information of various short-dated money market contracts (less than a year) and is set up to construct daily interest rate benchmark. An important side effect is that the MMSR data can be used to shed light on many other aspects of money market functioning. Relevant question could, for example, be how systemic or idiosyncratic stress affects the system.

Many data collections have become more granular allowing for made-to-measure aggregations. Whereas traditionally the task of classifying a particular data point was put upon the reporter, a current trend is to request information on a much more granular level and leave the classification to the authorities. For example, reporting firms no longer need to determine the sector of the counterparty but only need to supply the Legal Entity Identifier (LEI). The LEI, which will be discussed in more detail below, is a unique identification number for legal entities. Using the LEI, data can for instance be aggregated to a financial group. The LEI can also be used to generate sector aggregates. By leaving the aggregation process up to the data collector (i.e., the statistical agency, the supervisor or the central bank), consistent handling of the sector allocation is ensured. Furthermore, a change in the sector assignment of a particular firm can be applied consistently across all reported data.

[5] The Dodd-Frank Act in the US has similar stipulations.

A major advantage of data on a granular level is that it allows us to pinpoint the source of stress and the location of the most vulnerable parts of the financial system (Kwast *et al* (2010)). More granularity, however, also implies massive datasets and thus creates computational challenges. Moreover, if we are interested in feedback effects, we need to be able to trace such effects on a sufficiently high frequency. If the time scale is too coarse, we will only be able to see end results once they have already materialised. For example, if overlapping portfolios are an important driver of contagion because they open up a channel to transmit stress, then we need timely information of securities holdings. However, holdings are often, if collected at all on a granular level, only collected on a monthly or quarterly basis (see for instance mutual fund data in the US or the Securities Holdings Statistics in Europe). In the period in between, positions can move in many ways and in coming to stress test we need to make assumptions about how the portfolios change (Wang *et al* (2018)). In sum, examining the possibilities and challenges that new data and tooling bring in the stress test arena is useful, arguably not just for stress testing but for risk identification and management more broadly.

In stress testing many of these new developments come together. Stress tests often require the ad-hoc combination of data in order to be able to flexibly model unlikely scenarios. In the past, data availability dictated which scenarios were feasible to be stressed. For example, interbank exposures were only available as an aggregate without a split by counterpart. It was therefore very difficult to construct linkages from bank to bank. Without an accurate representation of the interbank network, tracing out the transmission of stress through the system is not possible.[6] Stress test were thus restricted to investigate how aggregate risk measures reacted to stress. Furthermore, modelling individual actors was often not possible since available information was not detailed enough to map out utility functions or supply and demand curves. Increased data availability might thus warrant a re-think of more traditional stress testing. A very promising avenue in this respect are heterogeneous agent models with network effects. These models are now computationally feasible and ready to deliver useful insights (Aymanns *et al* (2018)).

To be able to optimally conduct stress testing in this new environment, a number of challenges need to be resolved. First, our data should be of sufficient quality. In particular, the associated metadata should allow us to connect and combine datasets as needed. Second, a conceptual framework on how to handle the data needs to be adopted. Third, we need to develop the capability to tackle complex stress tests. This involves both an adequate IT infrastructure as well as a data-friendly company culture. We will discuss each of these points in turn. Then we will discuss a case study with the Securities Holdings Statistics (SHS) to show how granular data can provide the raw material for micro *and* macro stress tests.

Collecting data: national challenges and opportunities

Typically reporting requirements are developed in a `silo'. These silos are defined by the country specific legal frameworks that govern them. Mandates can be statistical, conduct-of-business (including market conduct), as well as micro- and macroprudential. Statistical

[6] See van Lelyveld and Liedorp (2006) and van Lelyveld *et al* (2011) for early efforts to construct network stress tests for the interbank and reinsurance markets, respectively.

mandates have the longest history. The English King William the Conqueror already ordered a stock-take of productive assets to determine the tax base as long ago as 1085. Market conduct mandates require data with the aim to ensure fair and efficient trading. Microprudential reporting is primarily aimed at collecting data to assess an individual firm while macroprudential supervision takes wider effects into account. For the latter type of supervision we are thus naturally more interested in data that can shed light on the interaction of different financial actors and the effects of financial markets on the wider economy. Each mandate has its own focus and incentives to coordinate across regulatory domains are limited.

Depending on a country's institutional framework, the legal mandates are assigned to different agencies potentially making sharing data difficult. Given the often complex governance, it is no wonder that reporting requirements are a patchwork of over- and underlapping components. Moreover, sharing data among agencies is often difficult because existing laws prohibit sharing confidential data. Even in the US, a country with a long tradition of publishing a wide range of financial information, this seems to be the case. This point was reiterated by Ruth Judson (Federal Reserve Board) at a recent G20 workshop on data sharing (Judson (2017)). She noted that "The U.S. has problems similar to those of other countries with regard to sharing data across agencies and countries".

Collecting data: cross-border challenges

Reporting and sharing data across national jurisdictions is even more complicated than in a specific jurisdiction. For example, the Basel Accords are a comprehensive set of rules for banks aiming to create a level playing field globally. The accords are themselves not legally binding, however. Basel committee member countries have to put the accord in national legislation first. For the current accord, Basel III, this is in Europe first and foremost the so-called CRD-IV package. CRD IV contains reporting requirements for licensed entities, that is banks. There are generally very few reporting requirements for firms outside the prudential regulatory perimeter. Furthermore, there is little coordination of reporting requirements for, for example, insurance undertakings and banks.

In the field of statistics, various bodies aim to coordinate regulation and the collection of (regulatory) statistics across jurisdictions. Standards are set by international institutions such as the UN, OECD, IMF and Worldbank. Agreements reached are then either implemented on a best-efforts basis or cemented in national law.

Collecting data: new opportunities through meta data and code distribution

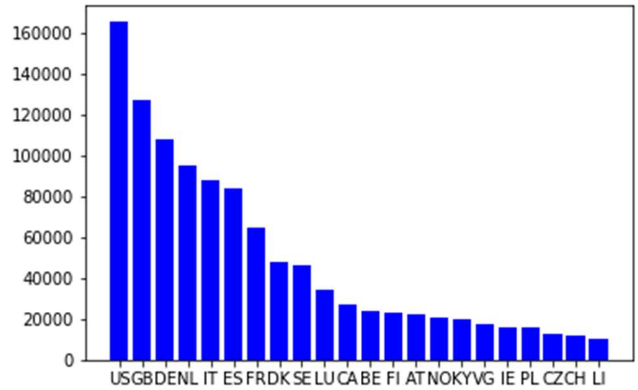
A consequence of how reporting requirements are typically designed, is that meta data is not well coordinated. Arguably, the precise reporting need to achieve their mandate is best determined by the relevant authorities. For example, microprudential banking supervisors are best placed to determine what information is needed to safeguard a bank. They should thus decide on what data points to request and at what frequency. Macroprudential supervisors might need other information because the scope of their mandate is wider. The same holds for other supervisors (e.g., insurance sector or financial markets supervisors).

Standardization across various reporting frameworks can be beneficial. Statistical agencies and standard setters have agreed on a long list of standards to ensure consistent measurement of social phenomena. For instance, there are International Organisation for Standardization (ISO) standards for country naming conventions. In financial markets, other conventions have been agreed on such as the International Securities Identification Number (ISIN, ISO 6166) and the Committee on Uniform Security Identification Procedures identification number (CUSIP) used in the US and Canada. These conventions do not yet have a complete coverage in terms of products or global acceptance but good progress is being made. For example, if regional breakdowns are reported using standard ISO country codes then we can easily combine information from, say, pension fund exposures to mutual fund investments in particular countries or regions. More complex issues arise if we want to unequivocally define economic activity across sectors: what constitutes a sector? What are the defining characteristics of a loan? If different sectors agree to use the same meta data frameworks, combining and validating data across different sectors can be simplified significantly. The choices in one domain can thus have an external effect and affect the usefulness and efficiency of reporting elsewhere. If for instance banking supervisors agree to universally adopt the LEI, other financial supervisors can follow and then matching exposures of banks with those of, for example, insurers becomes much easier. Berner and Judge (2019) support this line of thought and argue forcefully that developing standards jointly with industry increases social welfare.

An example of a key identifier in metadata is the Legal Entity Identifier (LEI).[7] The advantages of the LEI are discussed in more detail in Bottega and Powell (2011). The LEI is a 20-digit, alpha-numeric code based on a standard (i.e., ISO 17442). It connects to reference information that enables clear and unique identification of legal entities participating in financial transactions. The LEI initiative is a joint effort initiated by the Group of 20, the FSB and many regulators around the world. They have emphasised the need to make the LEI a broad public good. The LEI Regulatory Oversight Committee is at the highest level of LEI governance. Below this is the Global LEI Foundation, which brings together the information curated by local LEI granting organisations. These are typically credit registries and chambers of commerce. The number of LEIs is growing rapidly, fostered by new regulatory requirements, and now exceeds 1 million active identifiers (LEI ROC (2018)). The ECB estimates that the LEI covers at least securities with a total value of €95 trillion worldwide as of November 2017. This is a 25% increase since the end of January 2017. Although much has improved recently, some major challenges remain. For one, as shown in Figure 3, the coverage is far from even nor complete.

[7] Other initiatives in the ESCB to improve data collection are the Banks' Integrated Reporting Dictionary (BIRD), the Integrated Reporting Framework (IReF), the Register of Institutes and Affiliates Data (RIAD), and the Centralised Securities Data Base (CSDB).

Figure 3 Number of Legal Entity Identifiers (LEI)



Source: GLEIF (www.gleif.org). Parsing by Martien Lubberink

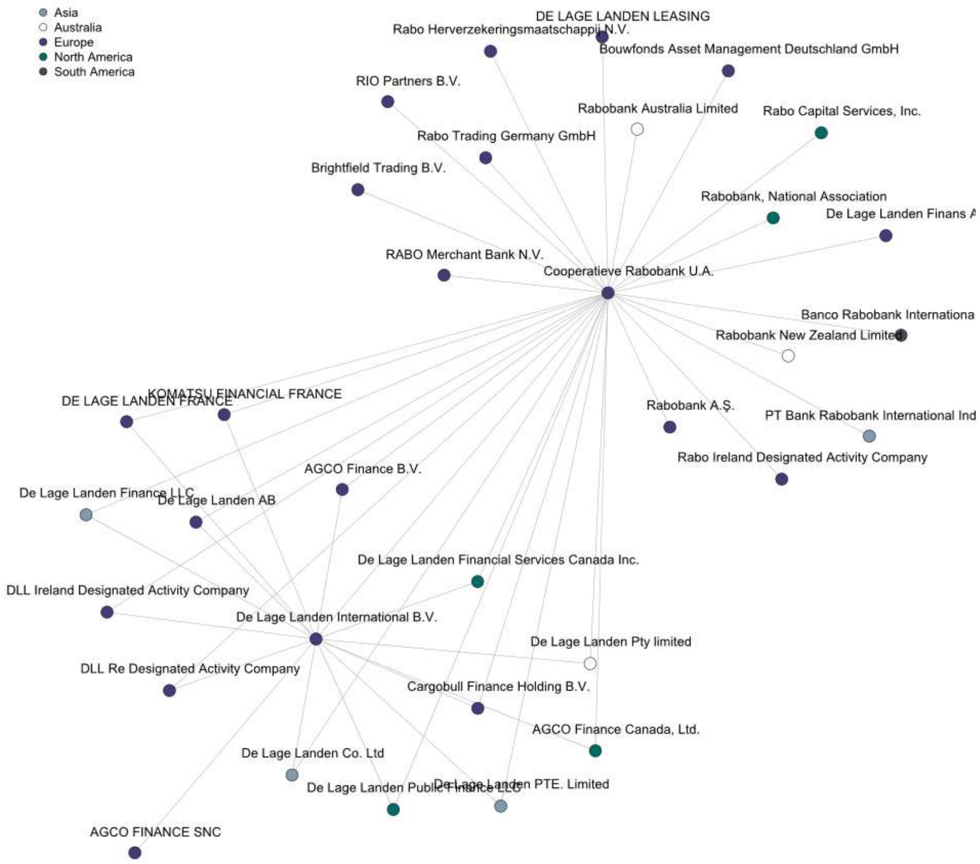
More technical challenges remain unresolved as well: for example, if a LEI lapses because the fees have not been paid, then a new LEI needs to be applied for. The entity will now have both an old inactive LEI as well as a new LEI. The old LEI might linger in reporting data for a long time if the entity entered into long running commitment such as a long-dated swap.

The LEI makes it possible to start assigning activity and risks exactly to those subsidiaries that are responsible. For example, for trading and holding of securities it is clear to whom to assign the flow and stock of securities. With multiple jurisdictions this allocation is of critical importance (Cerutti and Schmieder (2012), Fang and van Lelyveld (2014)). A poignant example is the default of Lehman where US authorities moved assets from Lehman's British subsidiaries before British authorities could react. Without information on the location of an asset (and how fungible it is) it is difficult to assess the robustness of a (local) institution.

To clarify how such allocation could work, we plot the legal structure of Rabobank in Figure 4. The information is taken from the GLEIF website and shows the structure and location of all the legal entities with LEIs. The different continents are coloured as explained in the legend. It is clear that even a relatively small player like Rabobank already has a very complicated structure spanning multiple jurisdictions. Note that this is only a partial picture since not all the units have a LEI attached to them. Given this structure, authorities in different jurisdictions can allocate activities to these units and assess whether these activities pose a local or global threat to financial stability.

Being able to break down a reporting entity into its constituent parts is very useful in stress testing because it allows for pinpointing where and how stress would affect a firm (Cerutti and Schmieder (2012)). If a bank for instance extends mortgages in two subsidiaries in two distinct geographic areas with quite different business cycles, then being able to stress the mortgage book using two different cycles can be quite valuable. Alternatively, if different resolution schemes are in force for different parts of a bank, claims and liabilities used in the stress test need to be broken down by jurisdiction (cf. Joint Forum (2012), in particular on the role of internal guarantees). Although a unique identifier such as the LEI is a necessary first step, we currently do not yet receive all the necessary information on a sufficiently disaggregated basis to allow for ad-hoc choice of aggregation. That is, a stress test of a bank with the option that some of the subsidiaries are allowed to default still requires extensive data work. In principle, however, we can construct a reasonably precise picture of a large and complex internationally active bank including a breakdown for the significant subsidiaries.

Figure 4 Legal structure of Rabobank



Ensuring a consistent reporting framework is quite a challenge although some progress is made. For example, the European Commission has undertaken an extensive study to assess whether the current reporting requirements are sufficiently harmonised (EC (2019)). Overall, the requirements are effective in meeting the objective of enabling supervisory authorities to fulfil their statutory tasks and mandates. While the EC concludes that EU-level requirements are coherent in a broad sense, they are found to have a range of inconsistencies both across as well as within reporting frameworks. Many of these inconsistencies appear minor and purely technical in nature. They nonetheless put a burden on reporting entities and supervisors. Key inconsistencies identified include non-aligned definitions, different formats of data fields or templates and inconsistent timing of mandated reporting. Note that EC's assessment did not back up the repeated claim by stakeholders that there are a significant number of duplicate requirements between different reporting frameworks. The EC did identify numerous cases of broadly or very similar data being requested. Authorities could thus consider if their information need merits multiple reports.

As an example of the mundane but in practice quite real complications that arise in combining information we can examine the differences for selected fields of the EMIR, MiFIR, REMIT and SFTR requirements as shown in Figure 5.[8] The data fields shown are present in all four frameworks but carry different labels. This could in principle be resolved by making a mapping based on the definitions. However, in some cases the definitions (or, equivalently, the admissible values) differ. For example, the data fields in the second row exhibit a validation inconsistency. Under EMIR, the status of the reported LEI shall be "Issued", "Pending transfer" or "Pending archival" while MiFIR reports are also accepted if the status of the LEI is '"Lapsed". For long-running swaps this might increasingly become an issue because lapsed LEIs would then generate a missing value in the EMIR data – but not in MiFIR. Likewise, there is a partial inconsistency in terms of content of the data fields in rows 2 and 3. Under REMIT, the reporting ID is not required to be an LEI308. Similarly, REMIT accepts a MIC, LEI, ACER code or ' XBIL' (for bilateral trades) as 'organised market place' while EMIR, MiFIR and SFTR always require this field to be populated with a MIC code. Linking REMIT to the other data sets is thus not straightforward.

Figure 5 Comparison of selected fields in EMIR, MiFIR, REMIT, and SFTR

EMIR	MiFIR	REMIT	SFTR
Field Identifier	Field Identifier	Field Identifier	Field Identifier
Currency of price	Price Currency	Price currency	Price currency
Report submitting entity ID	Submitting entity identification code	Reporting entity ID	Report submitting entity
Venue of execution	Venue	Organised market place	Trading venue

Source: EC (2019).

To overcome restrictions on sharing data, which could be used for more comprehensive stress tests, different approaches have been suggested – and in some cases executed as well. One way is to distribute code and only share the outcomes. Although the granular underlying data might be sensitive, the model outcomes – such as for instance regression coefficients – generally do not reveal the information encoded in the individual observations. In a project undertaken by the Basel Committee on Banking Supervision Working Group on Liquidity Stress Testing, participating researchers ran the code on their respective financial network data and then fed back the results to a shared repository. This resulted in an overview of network characteristics and network vulnerabilities for a wide range of networks (Anand *et al* (2018)).[9] A similar approach of code and result sharing is

[8] EMIR is a regulation on OTC derivatives, central counterparties and trade repositories, MiFIR covers reporting of transactions in financial instruments which are traded on a trading venues, REMIT deals with wholesale energy market integrity and transparency, and, the SFTR provides securities financing transactions reporting. See EC (2019) and the references and links therein for further details.

[9] The code for this project is available at https://github.com/imanvl/RTF_NTW_Horse. This approach is taken much further in the biomedical sciences where similar security and storage concerns are at play. See for instance the Global Alliance for Genomics and Health (https://www.ga4gh.org).

followed by the International Banking Research Network.[10] An alternative is to assign a trusted party who handles data collection and analysis.[11] For example, supervisors of the largest banks in the world have drawn up a legal framework that governs the exchange of very granular, bank specific information. This data is now being collected and analysed in the International Data Hub (IDH) hosted by the Bank for International Settlements.[12] The detailed reports are then sent back to the data contributors. Over the years the IDH has become more engaged in other regulatory workgroups, contributing to for instance analysis on the importance of CCPS, but external publication is still out of the question.

Data collection can have external effects. Although reporting requirements are almost always drafted with a single purpose in mind, the data collected can have multiple uses. For example, loan-level information on bank loans used in banks' internal credit risk models is generally primarily collected for microprudential supervision. With such data, supervisors can assess the validity and robustness of a bank's credit risk model. The data can also be used for other micro prudential purposes such as stress testing an individual bank. However, if such data is collected consistently *across* banks, then such data becomes potentially useful for macroprudential supervisors as well. The very same data can then also be used as input for an aggregate stress test. The aggregate stress test could in this case be micro-founded with attention for the distribution in credit risk and move beyond just examining credit aggregates. Alternatively, credit registry data – primarily collected for market transparency, statistical and monetary policy purposes – could be used for micro- and macroprudential stress tests. At DNB we are now investigating to what extent the information in the granular loan tapes, which we request ad-hoc in on-site microprudential examinations, overlaps with information in AnaCredit, the ESCB's credit registry. The advantage of AnaCredit is timeliness and cost. For AnaCredit, banks have put in place a regular reporting schedule which reduces the need for manual steps and hence costs. The disadvantage is that the definitions in place are not aligned with the supervisory definitions needed to assess compliance with, for example, Basel rules. At the very least, the AnaCredit information can provide a timely proxy for bank's credit risk.

To 'recycle' data in other domains might dictate additional requirements that do not seem immediately useful to the reporting parties. For example, the definition of what precisely constitutes a `bad' loan might currently be left up to a reporting bank (as the definition ties into internal control procedures) and the definition of when a loan is past due is then firm specific. This firm-specific choice might be determined by accounting rules, regulation or legacy IT systems. For understanding a bank's credit risk, the precise definition of past due is less important than that this definition is consistent over time and that the data is recorded accurately. To make the data useful for macroprudential analysis, however, it is important to use a single definition applied consistently across firms. Furthermore, if data is to be used for cross-country analyses then these definitions need to be agreed upon internationally as

[10] See https://www.newyorkfed.org/ibrn for more information.
[11] Yet another alternative is to encrypt the data in such a way that relevant information can be shared without revealing individual reporter's data. See Flood *et al* (2013).
[12] The IDH is part of the Financial Stability Board's Data Gaps initiative. See FSB (2009) for more details.

well.[13]

The ECB has undertaken a pilot to see how integration of highly granular data can be accomplished (Lauro en Traverso (2018)). The use case of this pilot was the analysis of the impact of the QE programme on banks' balance sheets on both the asset as well as the liability side. A key finding is that integrating two separate sets requires even higher data quality standards of the contributing sets than in the case of separate reporting; the data needs to be not just internally consistent but also match across reporting frameworks. Authorities thus need to be "prepared for integration" by putting in place integrated data management policies and adhering to best practices. Barbic *et al* (2017) provide detail on how, based on microprudential information, the ESCB has developed a data set with consolidated banking data which can be used for macroprudential analysis.

Data governance

To reap the benefits associated with combining different datasets on a regular basis we first need a conceptual framework of how we handle data. One possible starting point is the Four Quadrant Model we employ at DNB (Figure 6). In this model, the quadrants are delineated on two dimensions. First, on the y-axis we plot whether data is produced on demand (pull) or as part of a regular production cycle (push). An example of the former is the construction of a dataset for a dashboard built by a business analyst. The analyst could for instance be interested in a holistic view of the risk of a bank and would like to see risk metrics coming from different reports all presented in a single view. In this case, the analyst is actively asking for the data. An example of the pushed data is the regular prudential reporting process. The collection of such data is mandated by law and such data will be collected without the analyst actively asking for it. This is not to say the analyst is not interested, merely that he or she does not have to instigate the process. The second dimension, delineated on the x-axis, is whether the data collection is a structured, systematic process or an ad-hoc one. For instance, for producing macro-economic statistics, well-developed processes are in place which have been honed over the years. For ad-hoc research projects, in contrast, entirely new data sets or even methods to collect data need to be developed.

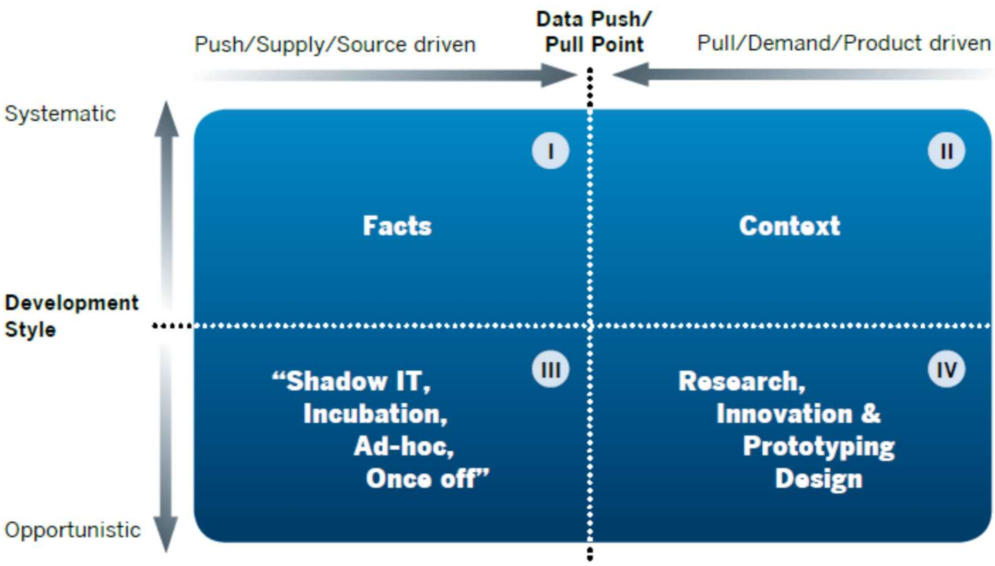
The combination of these two delineating dimensions results in four quadrants for which we require different data governance. To ensure that data in Quadrant I remains highly structured, we need to apply clear definitions and maintain high data quality. Generally, we define a Logical Data Model (LDM). An LDM is an exact definition of a data set. It defines aspects such as the number of variables, the storage type(s) and a description of what the variable supposedly measures. It also describes the relationships between variables. If reported data does not pass the data quality tests dictated by the LDM then one or more resubmissions are required until the data is of sufficient quality to be admitted to the production database. Quadrant 1 also requires an audit trail of data point changes otherwise the integrity of the data cannot be guaranteed. For example, analysist could – unintentionally or otherwise – change important benchmark interest rates or inflation rates.

[13] Initiatives such as the development of the Financial Industry Business Ontology (FIBO) are very helpful. FIBO is an open source business conceptual ontology of how financial instruments, business entities and financial processes work across the global financial industry (See https://edmcouncil.org/page/fiboproductsaccessre).

Quadrant II holds those (intermediate) datasets that users can, for instance, visualise in a dashboard. As discussed above, the analyst actively ask for data and are therefore also more involved in ensuring the quality of the data. Jumping ahead, Quadrant III covers those sets that require no centralised governance because they are not meant to be put onto the DNB data platform. Datasets in this quadrant are public or are local ad-hoc data collections. The ownership, and hence the responsibility of properly maintaining the data, lies elsewhere (for public data) or with the local user.

For stress testing, Quadrant IV is the most important quadrant. For the most part stress tests are bespoke exercises focusing on particular risks or scenarios. Although there might be a legal basis requiring periodic stress tests, the details are typically not spelled out. This implies the lower half of the y-axis of Figure 6. In some countries stress tests are becoming an annual fixture and hence the data requests might fit better in a Quadrant I process. However, to our knowledge, this has not happened anywhere. In terms of data governance, the best place for stress test analyses is Quadrant IV. Here the analyst can bring together various data sources usually required in stress testing. Since stress test scenario's change regularly, the input required will also differ each time.

Figure 6   The Four Quadrant Model for data governance



The choice of where to place a particular data set in the quadrant model depends on various factors. For example, the *importance* of the analysis using the data determines how much effort should be put into getting the collection process to the very highest level. With new data sources this presents a problem because the universe of possible useful analyses is not immediately apparent to management while the costs are real and immediate. *Repeated usage* might merit further investment in turning the reporting into a process. Another key factors is *available budget*. Finally, there is some value in the *flexibility required*. Tying yourself to the mast with a very prescriptive data intake process also implies that deviations from the process are costly. A deviation could for instance be requesting an accelerated delivery with lower data quality standards because a crisis is unfolding and authorities value noisy but timely information over perfect but later information. In such data files, some checks or required resubmissions might have been foregone. The information will be available faster but will also be noisier.

A major advantage of promoting a data set to Quadrant I is that data quality is unambiguous. This makes, for instance, combining data straightforward; since identifying keys are clear, no time is lost in matching data points from different sources. Furthermore, issues with missing or tainted should have been solved. It is up to the data owner to weigh the cost of promoting the data set to Quadrant I/II against the benefits of ease of use.

The challenge is to let the governance model create value in stress testing. As noted, stress testing often combines highly structured data (e.g., prudential CALL reports in the US or FINREP and COREP in the EU) with ad-hoc data sources (e.g., banks' proprietary model parameters). Moreover, if the stress test covers multiple sectors then information from different types of institutions (e.g. banks, insurers, and pension funds) needs to be shared by supervisors with different mandates. From Quadrant IV, the access to the high-quality structured data, which resides in Quadrant I, should be without impediments for those with the right access-rights. Given that the structure is clear, direct SQL-querying of databases can be implemented without impinging on confidentiality. Such direct access will reduce data wrangling costs, reduce errors and potentially make it possible to deliver stress tests results in a more timely fashion.

Note that the Quadrant model does not dictate the physical set up but primarily describes the governance of data. Traditionally, the physical setup (i.e., the hardware and the servers deployed on it) was tightly linked to the governance. Some servers were maintained in production with very tight governance while others were research and development servers with a much looser regime. With the appropriate access right procedures and rights maintenance software this tight coupling can become looser – applications and data can be anywhere because their use can be controlled.
The questions to ask about data governance can thus be decoupled from the physical set-up: What can we expect of the quality of the data? How important is traceability of data mutations? Who is allowed access? In practice, the Quadrant model still does map to different types of workspaces. At DNB, we use primarily the tooling for our large databases (SAS), a structured Business Intelligence (BI) process and strict governance for Quadrant I and II. For Quadrant IV we have a Research Area Network and a high-performance cloud environment where analysts have the freedom to deploy a much wider range of (open source) tooling. With this comes the responsibility to maintain this tooling and vouch for the solidity of the analysis.

In an ideal world all data would come with a completely fleshed out Logical Data Model (LDM). An LDM defines the structure expected in the data. It will stipulate data-types (e.g., string or float) and relations between data-points (for example, total assets should be equal to total liabilities). If the data complies with the LDM a complete description of the data is ensured. Such a description can for instance be used to ensure consistency across jurisdictions and communicate with reporting entities. It also points out exceptions but does not in itself provide the `business rules' that need to be applied to find practical solutions to reporting errors. For example, if the LDM prescribes using two-digit ISO country codes then a sensible business rule might be to also allow valid three-digit ISO codes even though strictly speaking these are a violation of the LDM. In the cleaning process a mapping from the two-digit to the three-digit ISO codes could be applied.

In practice, we define LDMs for a limited set of reports. Defining an LDM is costly because it entails fleshing out all the possible admissible permutations in a reporting framework. This is not trivial since many relevant regulations are ambiguous in their definitions (since more often than not they are the result of negotiations and the compromise has to reflect multiple views). Especially with ad-hoc requests there is insufficient time to work out a fully-fledged LDM. Moreover, it might not be cost-efficient to define a complete LDM for a single-shot data request.

Sometimes the LDM approach might seem overbearing and over-engineering for a single report. For instance, if we are designing an application to capture the contact information at supervised institutions (i.e., a mail address or phone number) it might seem overkill to work with an LDM. However, this contact-information might feed into a larger system that in the end is meant to be shared through an internal or a public register. If the data source is not captured unequivocally, then the register data will be tainted. In practice users across the institution will start to keep track of contact information leading to inconsistent and incomplete information. The costs materialise in the relatively simple address application while the benefits only materialise later for the organisation as a whole. The same holds for more important information such as the legal structure of supervised firms or the identifiers of financial instruments. Fairly attributing the costs of a local solution that is 'too much' at the local level but has institution-wide benefits is a significant management challenge.

The challenge is to find a balance between LDM fetishism and `anything goes' reporting. Although adhering to an LDM ensures that the data we admit to the final database meets a specified standard, there can be good reasons to diverge. For instance, if in a crisis situation a new datasource needs to be available on short notice, then defining an LDM will take too long. Alternatively, for small, one-off projects the costs of definition might create too much overhead. Nevertheless, the concept of an LDM is worthy to be applied more widely. The adoption of LDMs is however hampered because amongst policymakers – mostly economists and lawyers by training – the concept is relatively unknown. Unfortunately, LDMs are thus generally an afterthought to reporting requirements that have been the subject of long running negotiations. The structure and the content of the templates is then already set. Ambiguities that a rigorous approach such as an LDM definition process bring to the surface are then difficult to address. It would therefore be helpful if LDM-like thinking would be embraced from very early on in the reporting design. Furthermore, the external benefits which materialise only later and elsewhere in the organisation should be incorporated in weighing the pros and cons. One way to accomplish this is subsidise drawing up the LDM.

Acceptance of LDM-like thinking is hampered by, on the one hand, something we could term LDM-fetishism and, on the other hand, the typical way of how LDMs are presented to the non-initiated. LDM-fetishism is sticking to the model no matter what. This is not helpful as policy makers and researchers generally prefer noisy information over no information at all. In some, if not most cases, a greatly simplified LDM is more than adequate. The second reason LDMs are not widely embraced is because of the way they are presented. The typical end user is used to thinking in terms of a template or an Excel sheet. The LDM, however, reduces these Excel sheets to their essence and strips common elements to be defined in a separate place. The logical structure is then presented graphically. These wall sized graphics, printed in tiny fonts and to the uninitiated unintelligible symbols, are difficult to digest for most supervisors or policy officers. Modellers should put more effort in presenting the LDM in ways that appeal to the end user and that make clear how they add value.

Developing a stress testing capability

Once all these governance challenges, technical difficulties and conceptual issues have been resolved, we will be able to paint a complete picture of the financial ecosystem from all angles.[14] This requires a more or less comprehensive coverage of all relevant financial sectors. For some sectors, further detail is required to be able to identify systemically important entities. For example, data on lightly or non-regulated sectors is generally more sparse and if financial activity in these sectors picks up and starts to become a factor influencing systemic risk, then new ways of unearthing this information need to be found. For instance, although currently fintech firms' activity is only a fraction of the established financial sector (Frost (2020)), this might change in the future. For other sectors, a coarser aggregate can be sufficient. For instance, if contracts are linear then stress test effects are less susceptible to tail scenarios and we thus have less need to the full distribution of exposures. It also requires that we collect stocks of all relevant balance sheet positions (including off balance sheet items and derivatives exposures). Furthermore, to detect fast moving risk shifting we will need to collect financial market transaction data (i.e., flow data).

Ideally, the validity of the data is beyond reproach. In practical stress testing, significant effort is spent on discussions about whether the often patchy data quality warrants the conclusion. However, discussions about the interpretation of stress test results should ideally not be about the validity of the bare facts; these discussions should focus on the conclusions and their policy implications. In practice, some of the data might indeed be noisy and – as some argue – this is inevitable. In stress test we often have no choice, however. We should be pragmatic and get the data up acceptable levels with sensible business rules.

With data on the entire ecosystem, we will be able to deliver the raw material for both micro- and macroprudential stress tests. In a sense, we can seamlessly zoom from a highly aggregate macro level all the way down to individual transactions of individual market participants. Moreover, we can also connect markets and see what market participants or sectors are doing in different instruments. Depending on the mandate, regulators and supervisors are interested in different scales of aggregation. For instance, a conduct-of-business supervisor will be interested in individuals trading with insider knowledge. A microprudential supervisor will be looking at the solvency of a particular firm, while a macroprudential authority will look for vulnerabilities at the macro level. With granular data, views for each of these can be generated from a single data source. To be able to zoom into different time scales we need the most granular level to start with.

To illustrate the notion that collecting granular data allows for flexible "zooming", we show data on securities holdings in the Euro area. For compiling national accounts

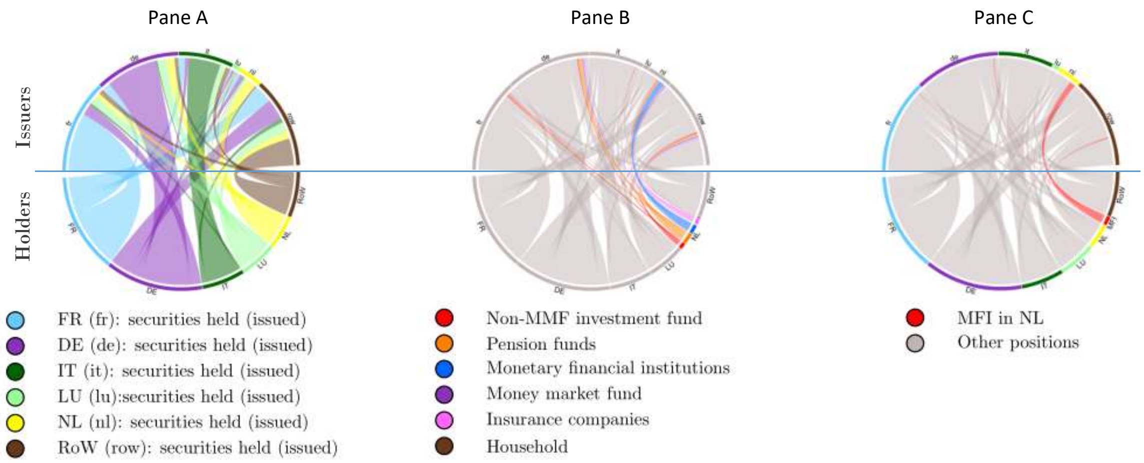
[14] See Bijlsma *et al* (2018) for an example.

statistics, ESCB member banks collect securities holdings of Euro area residents combined in the Securities Holding Statistics (SHS). Each central bank collects individual securities holdings from relevant entities in their jurisdiction. In the Netherlands, the reporting sample includes around 800 reporters. In principle, we can thus move from country aggregates all the way down to individual firms holding a particular instrument. We visualise this process in Figure 7 in three panes.[15] In each pane we use the same data but focus on different levels of aggregation. For clarity, we collapse smaller countries to a 'rest of the world' (RoW) category and show holder countries in the bottom half of the circle while issuers are shown in the top half.

In the left pane (Pane A), we show the total exposures of country holders to various issuing countries. This could be called the `macro' view. Here we see that France is the largest holder (8.3 Tr euro) and that its holdings have a significant home bias: 60% of its holdings are issued domestically with 500 Bn euro (6%) issued in Germany a distant second. It is clear that home bias is an entrenched stylised fact. This level of aggregation is useful for macro-economic analyses of international capital stock and flows and is most useful for macro stress tests.

Next, if we are interested in a particular sector, we can zoom in as in Pane B. We term this the `meso' view. Here, we show the exposures of the different sectors in the Netherlands to different other issuing countries. The largest holdings are concentrated in the pension fund sector, with the Monetary Financial Institutions (MFI) or banking sector a close second. We see that for most sector the largest exposure is to Dutch firms (i.e., a self-loop).

Figure 7 Securities data at different levels of granularity



The rightmost pane (Pane C) highlights the exposures of a single sector, in this case the banking sector. Showing all sectors, as in Pane B, can sometimes conceal important exposures if we are interested in just a single sector. Alternatively, we could use the colour to denote the issuing country. Such a view would allow us to focus on the source of foreign shocks.

Zooming in even further – not shown due to data confidentiality – we get to the `micro'

[15] See Sigl-Gloeckner (2018) for a survey of network visualization methods, in particular for the SHS data.

view where we could show to what country-sector one particular pension fund is exposed. This would be of interest to the pension fund's supervisor when assessing a fund's country risk. Finally, we can discern the 'nano' view. Here we could highlight the exposure of a single fund to a single instrument (identified by ISIN). As an additional flourish, we can show all related instruments issued by the same corporate group or, in the same vein, to a particular group of instruments (cf Boermans and van Wijnbergen (2018)).

Having the ability to zoom is potentially extremely valuable for the efficiency of stress testing. The same source data can be used for stress testing from different micro and macro angles. For example, if we are worried about country risk exposures of a particular sector, we take a `meso' view and stress the valuation of the instruments held. This is generally a financial stability concern. Conversely, if we are interested in a micro-prudential approach, we can just select the individual institution and apply the stress parameters to its portfolio items. The same data can thus serve different constituencies. Incidentally, having granular data at our disposal also means that macroprudential stress tests can have micro-foundations, meaning that they can model/simulate shock amplification with much higher degrees of verisimilitude, potentially capturing otherwise-hidden risks.

With the underlying data of the SHS, national authorities can also move to the individual firm, be it a bank, a pension fund, or an insurance firm. The portfolios of the larger banking firms are available in full detail. On this level, we can currently already add a lot of other firm-level data. For example, we can add public information from commercial data providers such as SNL and ORBIS. For research purposes, we can add prudential information from *inter alia* FINREP and COREP. With such information we can for instance study how a bank suffering liquidity stress – as measured in the regulatory liquidity coverage ratio (LCR) reports – reacts by engaging in precautionary hoarding of liquid securities (Acharya and Merrouche (2013)) or is forced to fire sale (Diamond *et al* (2011)).

Going forward, it should also become much easier to zoom using other cross-country data collections. For example, the European Market Infrastructure Regulation (EMIR), Securities Financing Transaction Regulation (SFTR) and Money Market Statistical Reporting (MMSR) are all granular data collections that can be analysed on different aggregation levels.

Consistent data across different markets also allows for stress testing of connected markets. For example, a recent trend is that exposures in derivative market are ever more mitigated by the exchange of collateral. This is driven by regulation mandating central clearing but also by firm's reduced appetite for counterparty credit risk. By exchanging collateral, counterparty credit risk is reduced significantly. However, since collateral requirements are very sensitive to market volatility, the reduction in credit risk is accompanied by an increase in liquidity risk. Clearly, sudden spikes in market volatility translate directly into collateral calls. And funding such calls can be particularly expensive or difficult at precisely these same times. To properly understand either the derivative or funding markets such as the repo market, we should thus be able to analyse the two markets jointly (Adrian *et al* (2013)).

Collecting consistent information for an entire market might allow for better risk assessment than what financial markets can currently achieve. Financial markets are generally very efficient in digesting information. However, a financial market cannot price a risk it cannot observe. In many OTC markets, only a few core players can form a reasonably coherent view on all participants' positions. Such market-making parties can capitalise on the information embedded in order flow or – accumulated over time – knowledge of counterparties' positions (Duffie and Manso (2007)). In a dispersed market, however, such centralised knowledge is not available. The market can thus sometimes miss pockets of risks due to, for example, concentration risk. A well-known example is AIG, the American insurer that turned out to be a key player connecting the financial markets in the 2008 crisis. Information on the network as a whole can potentially improve on fragmented information. For example, Squartini, Garlaschelli and Van Lelyveld (2010) show that network information can show fragmentation in interbank network connections significantly earlier than market signals. Such market fragmentation can potentially hamper a market's ability to redistribute risks.

Note that stress tests are often kept simple on purpose in order to be able to tell causal stories (Drehmann (2009)). Some data analytics methods do not fit well with such a strategy. The results from machine learning algorithms, for instance, are difficult to trace back to the actions of individual actors. This makes explaining the outcomes to mostly uninitiated users of the results (i.e., management or the general public) rather difficult. Without a comprehensible narrative, defending stress test outcomes becomes very difficult. One way forward would be to try to model the complexities in the tradition of Herbert Simon (Simon (1962)). The understanding of the underlying processes would allow for the translation into causal stories. Another way forward would be to tell a broader story about how shocks to the system, depending on their nature and magnitude, can either be absorbed or are destabilizing (Wiersema *et al* (2020)). Finally, in some cases contributions to particular outcomes can be attributed to particular agents or risk drivers using the concept of Shapley values (Joseph (2019)).

Challenges in a new environment

The new possibilities for stress testing discussed so far also come with several challenges as well in terms of a) governance, b) the IT capacity to handle the data volume, and c) the skill sets needed. We will discuss each of these in turn.

Governance. Collecting data on a more granular level implies more, often sensitive, detail. Moreover, the granular level sometimes contains information on natural persons. For example, loan level information might contain natural persons' names which are in principle not needed for stress testing (beyond being the primary key to link other information (e.g., tax records). Handling such data demands additional care. In this case, it might be useful to anonymise the loan level data while preserving the ability to combine the data with other sources. In Europe, the General Data Protection Regulation (GDPR), which is being enforced as of May 2018, now comes with much tighter regulation of what firms and authorities can and cannot do with natural persons' data. In collecting data that contains or might contain natural persons data, adequate thought needs to be given to ensure fair use while still adhering to the GDPR.

Volume. To collect, digest and store larger volumes of data, authorities might need to explore unfamiliar terrain, for example storing data outside their own servers. Although generally the costs of commercially available cloud services (i.e., computing and storage), have come down drastically, many government agencies are not yet ready to take advantage of this since this would mean outsourcing IT capacity in the cloud. The decision to outsource is an important strategic choice that also comes with operational, security and legal risks. Arguably, security risks have economies of scale and could thus possibly be better handled by larger and more specialised outside parties. Note that in some cases, the data is not allowed to leave the jurisdiction and in smaller jurisdictions this limits the number of providers that are able to provide the services demanded.

To get a better understanding of the challenges, possibilities and limitations of using cloud services, DNB has started an initiative in early 2018 to bring confidential data to the cloud. Here, we can more easily apply advanced and computationally expensive techniques to datasets of several billion observations. Such volumes are already too big for most central banks and supervisors and thus potentially useful insights are foregone. Moreover, one might even argue that the mandate to collect transaction-level data is rendered pointless if analysts do not have the IT infrastructure needed to study the data. One of the pilot projects undertaken to build on our capacity to analyse at scale is a computationally expensive stress test of margining requirements in the OTC interest rate derivatives market. The stress tests looks into the liquidity risk of margin procyclicality which has already attracted some attention in the academic literature (cf Glasserman and Wu (2018)).

Skill set. If an organisation is to become truly data driven, then data should be embedded centrally in the business process. This requires that those involved should not be ‘data challenged’. Some would argue that regulators are more likely to be ‘data challenged’ than the regulated, and that this creates its own pitfalls. Traditionally, market conduct supervisors are trained lawyers while central banks rely on economists. Both lawyers and economists generally have attractive outside options in the private sector. Authorities cannot remunerate on the same level as the private sector but try to compete by providing a better work-life balance and an appeal to civic duty. In addition, working in government for some time allows professionals to build a network and get a much better understanding of the regulatory process. This is very valuable for a career in the private sector and the wages foregone can thus be seen as an investment in higher future income in the private sector. However, Luca Enriques has argued that such a trade-off might not be relevant for IT engineers and Data Scientist in Regtech or Suptech.[16] Developing IT applications for regulatory compliance or supervisors is much less dependent on soft skills and localised knowledge. It will therefore be more difficult for government agencies to attract top-level IT engineers. Given the ample outside opportunities for people with these skills, it becomes even more important to offer interesting and rewarding jobs. Hopefully, the very rich data sets central banks and supervisors can use and the appeal of a meaningful job that offers rewards to society at large is sufficiently convincing to data scientists and IT engineers to accept a pay cut.

It is crucially important that data scientists interact with supervisors and policy officers. In doing so, data scientists will be inspired by real world problems, whereas their colleagues in policy and supervision will get a better understanding of where central banking and

16

supervision can benefit in practice from the data revolution. New techniques such as artificial intelligence and machine learning will not drive out professional judgement, which in terms of analysing shocks and their impact on the financial sector will remain essential for figuring out which shocks are feasible and for thinking outside the box more broadly. In terms of policy making on the basis of stress testing, judgement calls will remain essential, for instance in case of dilemmas where a choice is needed between two options which each affect a different group in society.

Discussion and conclusion

To really make new opportunities in data pay off for stress testing requires hard work. The new and enticing possibilities are not a free lunch. To achieve them means to adhere to data governance principles, for instance by using meta data consistently and on a large scale, collecting and cleaning new data sets, applying LEI and UPI codes etc. For a single stress testing project the costs may seem disproportional. But usually there are substantial external benefits. As we go further down this road, new opportunities are emerging for an all-encompassing risk management approach on the basis of modelling, in a comprehensive way, the financial ecosystem.

In this chapter we have explored the opportunities that more granular data and closer to real time reporting bring. Granular – even transaction level - data allows us to aggregate the data in many ways. This is ideal for stress testing as here flexibility is key; the strength of stress testing is that it is possible to entertain unlikely scenarios that have never materialised. Often these involve new markets or players for which we don't regularly collect data. These opportunities also come with challenges. Data collection is often organised in silos. For instance, insurance supervisors focus on insurance firms (e.g., Solvency II) while reporting for monetary policy operations (e.g., MMSR) covers primarily banks. In looking at the impact of interest rate setting on the wider economy it might, however, be interesting to analyse how money market funding conditions are affected by insurance firms' repo transactions – and vice versa. Such analysis is hampered by inadequate coordination on reporting standards (frequency, metadata, etc.).

Having more and more granular data in-house also puts more weight on the governance of that data. For instance, who is the owner of the data? Dispersed ownership is an invitation to free riding and cost shifting. At DNB, the owner is therefore always a single person, although he or she is free to involve others for input or funding. Given that more granular data is often also useful more broadly, finding an organisational form that takes into account external effects is becoming more important. Furthermore, the owner should aim for maximally distributing the data to be able to realise its full potential. This implies a 'share, unless …'-policy which should be accompanied with well implemented access rights on the ground. Such a data policy should hold within organisation as well as for the interaction between organisations involved – that is the macroprudential supervisor (ESRB and OFR/FSOC in the EU and the US, respectively) and supervisory authorities and central banks.

The flood of new data potentially also has reputational risk for authorities. For some parts of the system authorities knew very little and were thus blind sighted for the build-up of risks. In the aftermath of the crisis, a massive increase in data reported ensued. The challenge authorities now face is to effectively use all the data and turn it into information and actionable analysis. Failing to incorporate the new data into analyses runs the risks of instigating a large public backlash the next time a crisis materialises. Some might conclude that authorities did have crucial information reported but failed to act on it.

In sum, new data and tooling have the potential to significantly improve stress testing but challenges remain. The sheer volume of data requires an investment in data processing capacity. More importantly, this requires a change in the way we use the data. Traditionally, those involved in stress testing have a supervisory or a financial stability background. Knowledge of database management and optimizing database queries is thus often insufficient to handle large volumes efficiently. Another challenge is how to communicate results to senior management and lay people. Scrolling through the data in an Excel sheet to get a feel for the data is no longer feasible. In presenting results, more care should thus be given to supply appropriate dashboarding as well. Characterizing high dimensional data with new concepts – such as the eigenvector centrality of a bank in a market – needs to be accompanied with sufficient explanation to instil a sense of comfort in decision makers.

References

Abad, J., Aldasoro, I., Aymanns, C., D'Errico, M., Fache Rousová, L., Hoffmann, P., Langfield, S., Neychev, M., Roukny, T., 2016. Shedding Light on Dark Markets: First Insights from the New EU-wide OTC Derivatives Dataset. ESRB Occasional Paper 11.

Acharya, V. V., Merrouche, O., 2013. Precautionary Hoarding of Liquidity and Inter-Bank Markets: Evidence from the Sub-prime Crisis. Review of Finance 17, 107-160.

Adrian, T., Begalle, B., Copeland, A., Martin, A., 2013. Repo and Securities Lending. Federal Reserve Bank of New York Staff Reports 529.

Anand, K., van Lelyveld, I., Banai, A., Christiano Silva, T., Friedrich, S., Garratt, R., Halaj, G., Hansen, I., Howell, B., Lee, H., Martínez Jaramillo, S., Molina-Borboa, J. L., Nobili, S., Rajan, S., Rubens Stancato de Souza, S., Salakhova, D., Silvestri, L., 2018. The Missing Links: A Global Study on Uncovering Financial Network Structure from Partial Data. Journal of Financial Stability 35, 107-119.

Aymanns, C., Georg, C.-P., Golub, B., 2018. Illiquidity Spirals in Coupled Over-the-Counter Markets.

Barbic, G., Borgioli, S., Klacso, J., 2017. The journey from micro supervisory data to aggregate macroprudential statistics. ECB Statistics Paper Series 20.

Berner, R., Judge, K., 2019. The Data Standardization Challenge. ECGI Working Paper Series in Law 598.

Bijlsma, M., Castro Campos, M., Chaudron, R., Jansen, D.-J., 2018. Building a Multilayer Macro-Network for the Netherlands: A New Way of Looking at Financial Accounts and International Investment Position Data. IFC Working Papers.

Boermans, M. A., van Wijnbergen, S., 2018. Contingent convertible bonds: Who invests in European CoCos? Applied Economics Letters 25, 234-238.

Bottega, J. A., Powell, L. F., 2011. Creating a Linchpin for Financial Data: Toward a Universal Legal Entity Identifier. Finance and Economics Discussion Series 7.

Cerutti, E., Schmieder, C., 2012. The Need for "Un-consolidating" Consolidated Banks' Stress Tests. IMF Working Paper WP/12/288.

Diamond, D. W., Rajan, R. G., 2011. Fear of Fire Sales, Illiquidity Seeking, and Credit Freezes. The Quarterly Journal of Economics 126, 557-591.

Drehmann, M., 2009. Macroeconomic stress-testing banks: a survey of methodologies. In: Quagliariello, M. (ed.), Stress-testing the Banking System: Methodologies and Applications, Cambridge University Press.

Duffie, D., Manso, G., 2007. Information Percolation in Large Markets. American Economic Review 97, 203-209.

Ehlers, T., Kong, S., Zhu, F., 2018. Mapping Shadow Banking in China: Structure and Dynamics. BIS Working Paper 701.

European Commission, 2019. Fitness Check of EU Supervisory Reporting Requirements. Commission Staff Working Document SWD.

Fang, Y., van Lelyveld, I., 2014. Geographic Diversification in Banking. Journal of Financial Stability 15, 172-181.

Financial Stability Board, International Monetary Fund, 2009. The Financial Crisis and Information Gaps. Report to the G-20 Finance Ministers and Central Bank Governors.

Financial Stability Board, International Monetary Fund, 2017. Second Phase of the G-20 Data Gaps Initiative Second Progress Report. Report to the G-20 Finance Ministers and Central Bank Governors.

Flood, M., Katz, J., Ong, S., Smith, A., 2013. Cryptography and the Economics of Supervisory Information: Balancing Transparency and Confidentiality. OFR Working Paper 11.

Foster, W. T., 1922. The Circuit Flow of Money. American Economic Review 12, 460-473.

Frost, J., 2020. The Economic Forces Driving FinTech Adoption across Countries. DNB Working Paper 663.

Glasserman, P., Wu, Q., 2018. Procyclicality in Sensitivity-Based Margin Requirements. Working Paper.

Joseph, A., 2019. Shapley regressions: A framework for statistical inference on machine learning models. XArchiv Working Paper.

Joshua Graff, 2018. The real shape of Britains post-Brexit talent brand. Linkedin Pulse.

Judson, R., 2017. Enhancing Data Availability: Recent U.S. Experience with Banking Data. In: Workshop on Data Sharing, Deutsche Bundesbank.

Kienecker, K., Sedlacek, G., Turner, J., 2018. Managing the processing chain from banks source data to statistical and regulatory reports in Austria. OeNB Statistiken 3.

Kwast, M. L., Holden, S., Jurcevic, D., Van Lelyveld, I., 2010. Norges Bank Stress Testing of Credit Risks - Report of an External Review Panel.

Lauro, B., Traverso, R., 2018. Data Fitness for Integration. Mimeo.

Legal Entity Identifier Regulatory Oversight Committee (LEI ROC), 2018. The Global LEI System and regulatory uses of the LEI. Progress Report.

Levels, A., De Sousa Van Stralen, R., Petrescu, S. K., Van Lelyveld, I., 2018. CDS market structure and risk flows: the Dutch case. DNB Working Paper 592.

Sigl-Gloeckner, P., 2018. Visualising Financial Systems. Thesis Imperial College London.

Simon, H. A., 1962. The Architecture of Complexity. Proceedings of the American Philosophical Society 106.

Squartini, T., van Lelyveld, I., Garlaschelli, D., 2013. Early-Warning Signals of Topological Collapse in Interbank Networks. Nature Scientific Reports 3, 3357.

The Joint Forum, 2012. Report on intra-group support measures.

van Lelyveld, I., Liedorp, F., 2006. Interbank Contagion in the Dutch Banking Sector: A Sensitivity Analysis. International Journal of Central Banking 31, 99-133.

van Lelyveld, I., Liedorp, F., Kampman, M., 2011. An empirical assessment of reinsurance risk. Journal of Financial Stability 7, 191–203.

Wang, D., van Lelyveld, I., Schaumburg, J., 2018. Do Information Contagion and Business Model Similarities Explain Bank Credit Risk Commonalities? DNB Working Paper 619.

Wiersema, G., Kleinnijenhuis, A. M., Wetzer, T., Farmer, J. D., 2020. Inherent Instability: Scenario-Free Analysis of Financial Systems with Interacting Contagion Channels. Institute for New Economic Thinking 10

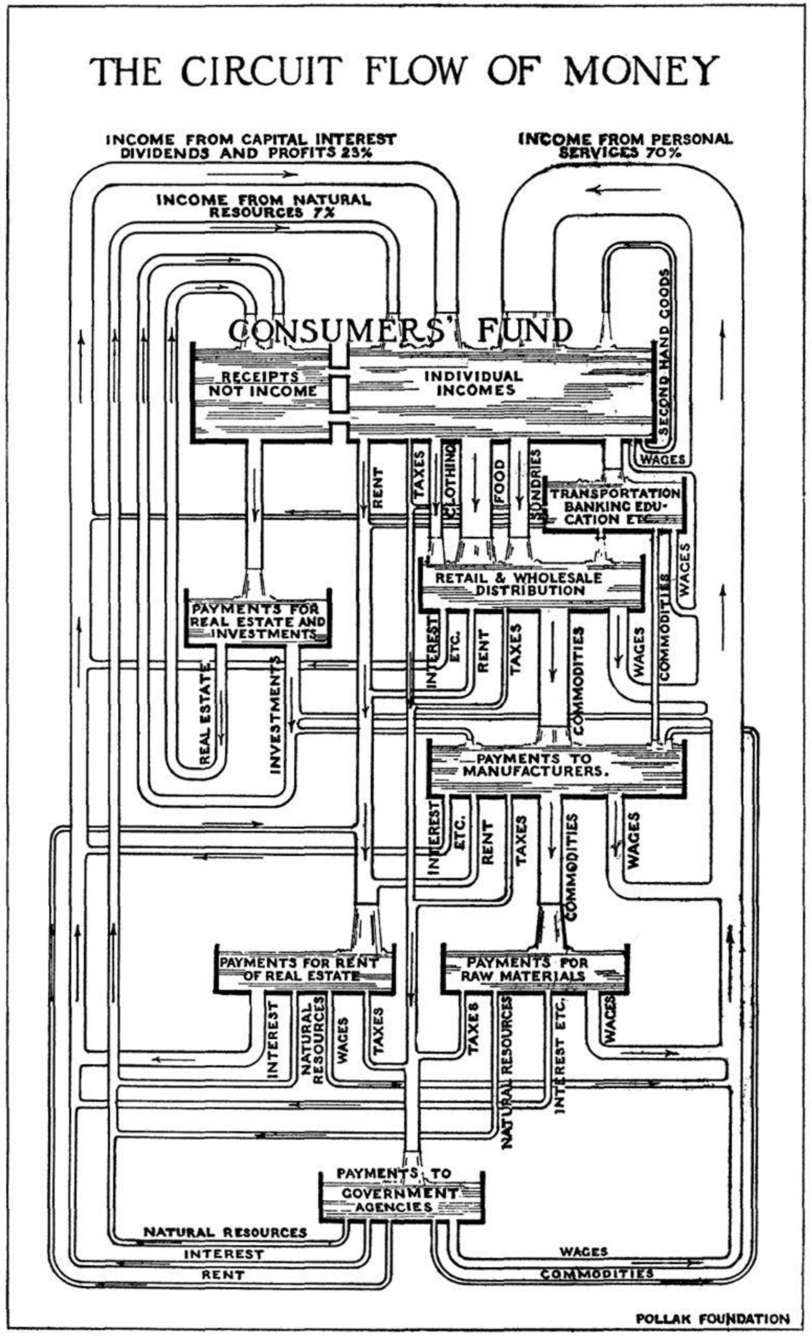
Granular data and stress testing: stepping up to the challenge

IFC-BNM-ECB Satellite on ***"Post-crisis data landscape: micro data for the macro world"***
Kuala Lumpur, Malaysia, 16 August 2019

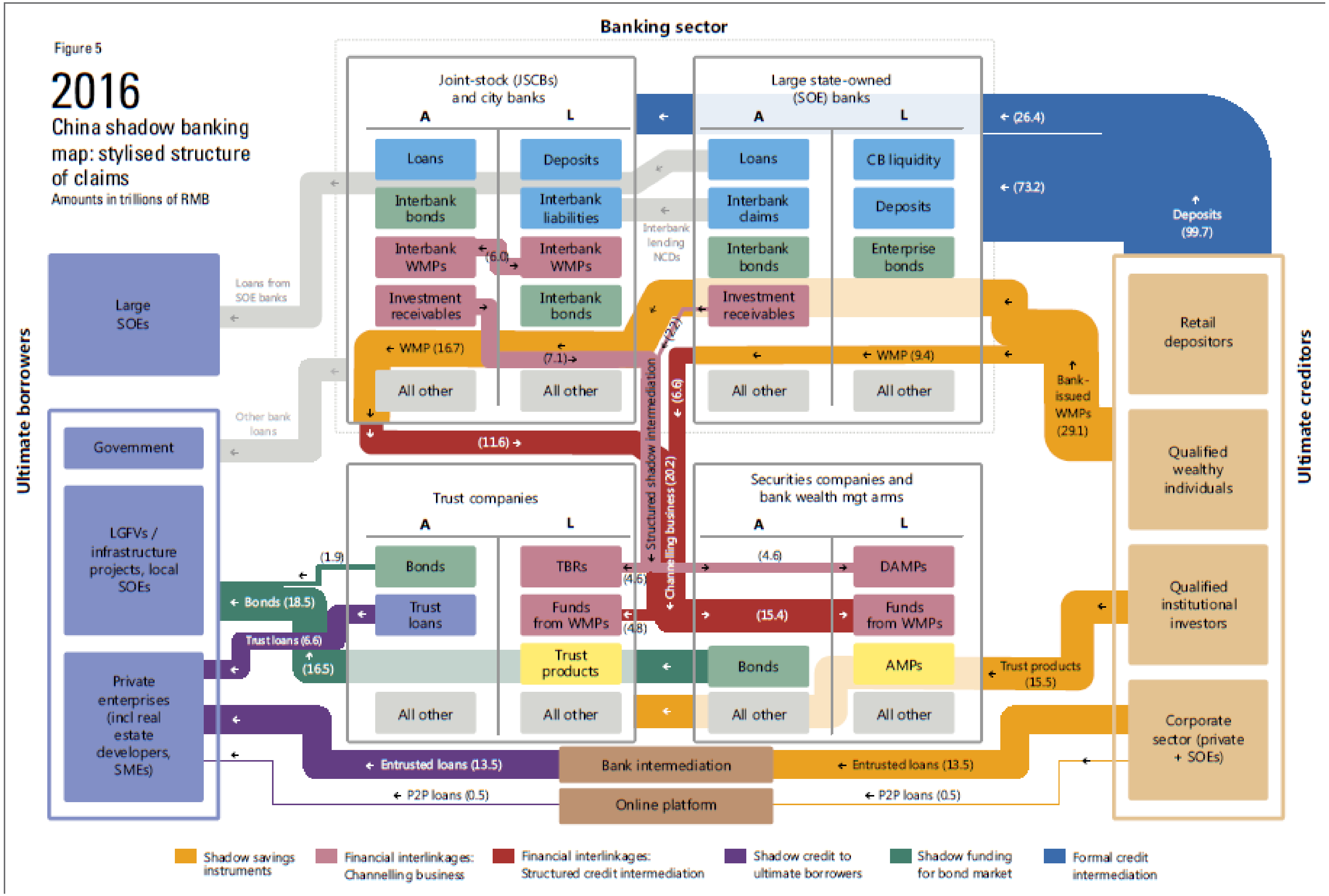
Cees Ullersma

DeNederlandscheBank
EUROSYSTEEM

The usual disclaimer applies



Source: Foster (1922), “The Circuit Flow of Money”, AER



Source: Ehlers et al (2018), “Mapping shadow banking in China: structure and dynamics”, BIS WP

Main message:

- Granular data: allows for a mapping of financial eco system
- Data on stocks (balance sheets) and flows (transactions) available more widely
- New opportunities for analysis of shocks
  - Identifying troubled firm, sectors, and markets
  - Trace propagation of shocks through system
- Very promising for stress testing

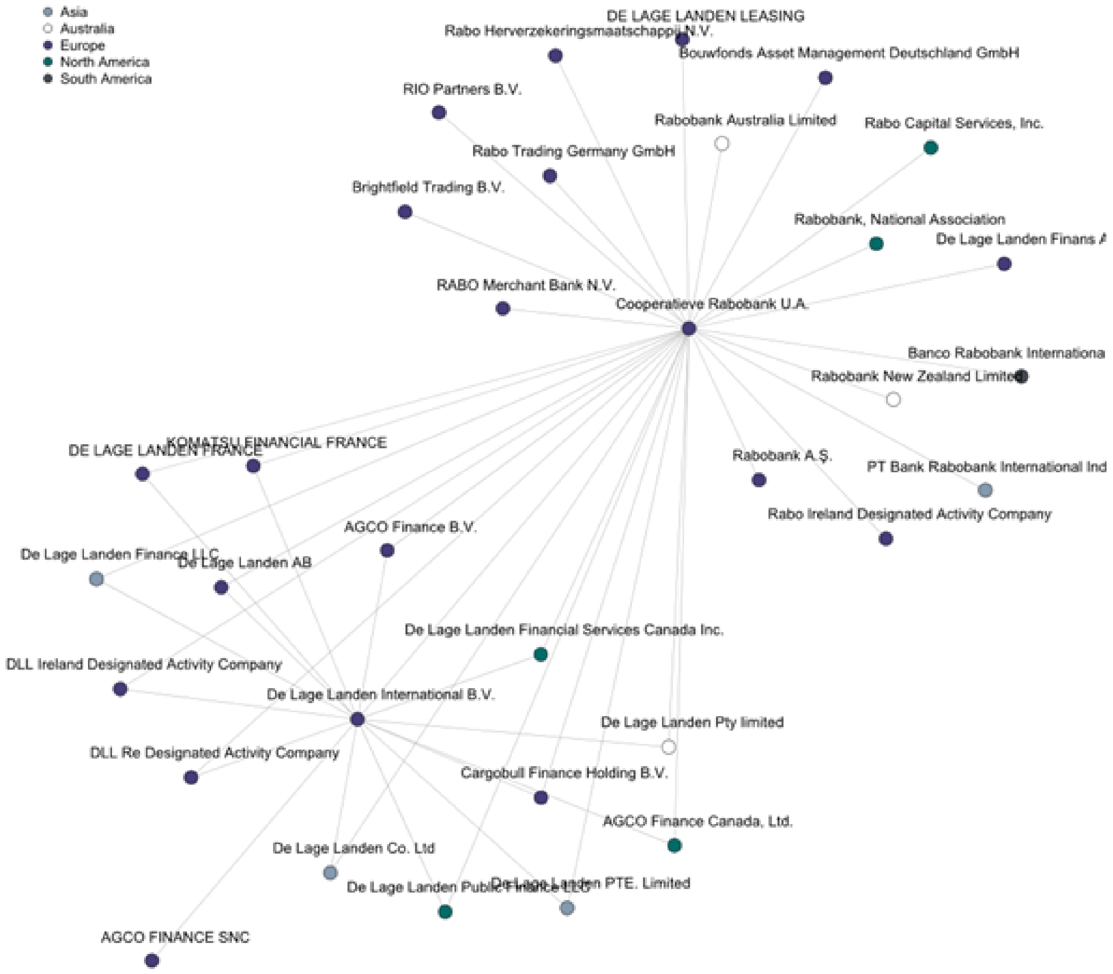
Excellent starting position

- Availability of granular data
  - Securities Holding Statistics (SHS)
  - Data Gaps → International Data Hub at the BIS
- New OTC reporting
  - Derivatives (EMIR)
  - Money markets (MMSR)
  - Repo (SFT)
- Key competence of central banks and supervisors

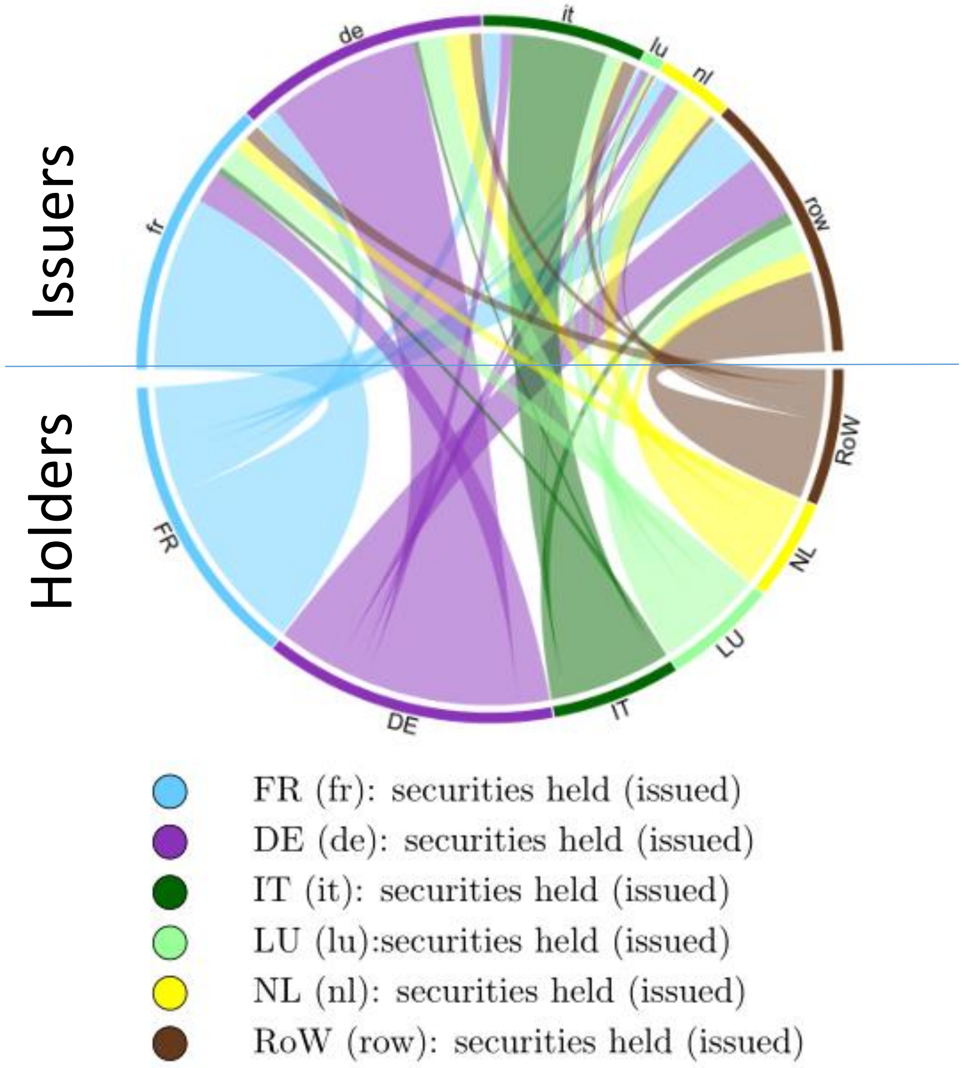
Challenges and the way forward

- Data sources diverse
  - Structured & unstructured
  - Numeric & textual
  - Inside & outside
- Obstacles to data sharing
  - Legal restrictions within and across jurisdictions
- Adequate data governance
  - Data ownership and access should be well anchored
- Have the basics in place
  - data catalog, meta data, LEI/UPI
- Open attitude to alternative ways of cooperating
  - Sharing files (eg. mailing code, Sharepoint)
  - GIT: most widely used version control system
  - Container technology

The importance of the Legal Entity Identifier (LEI)

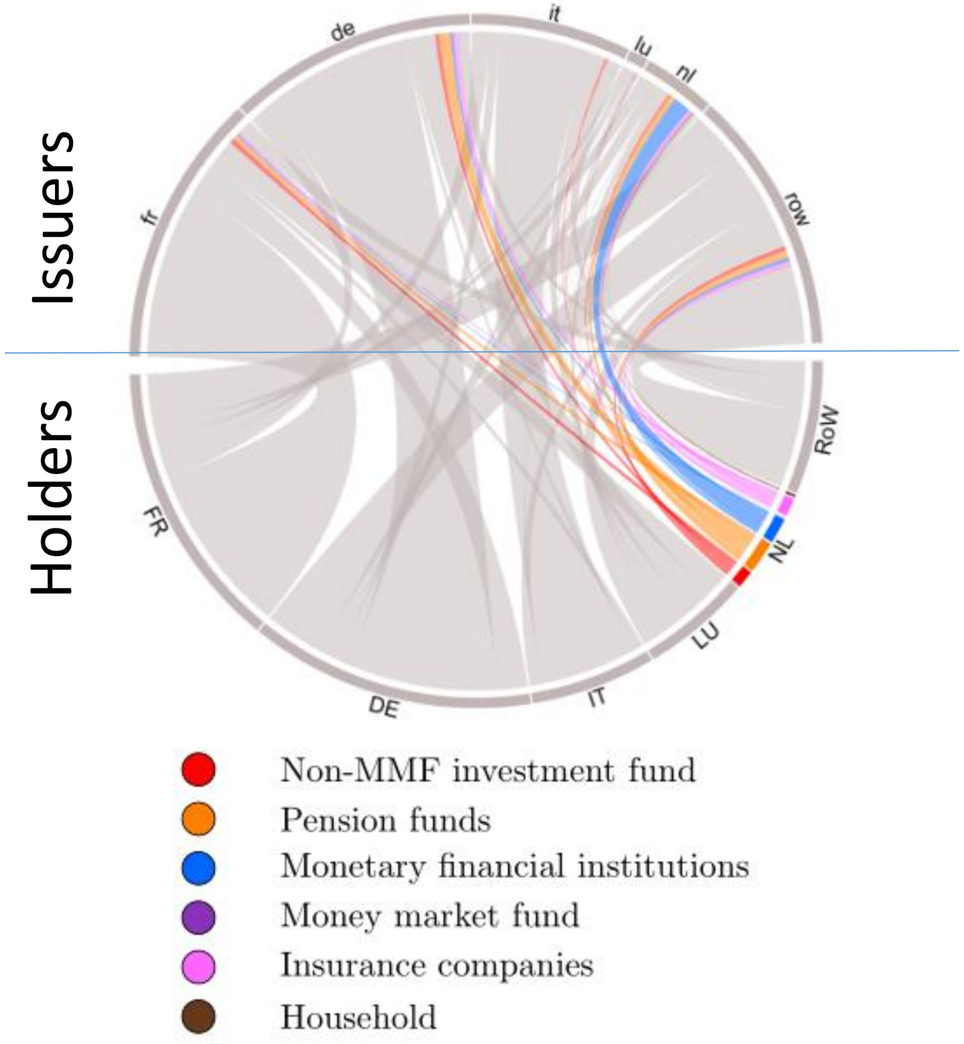


The same granular data can serve multiple goals



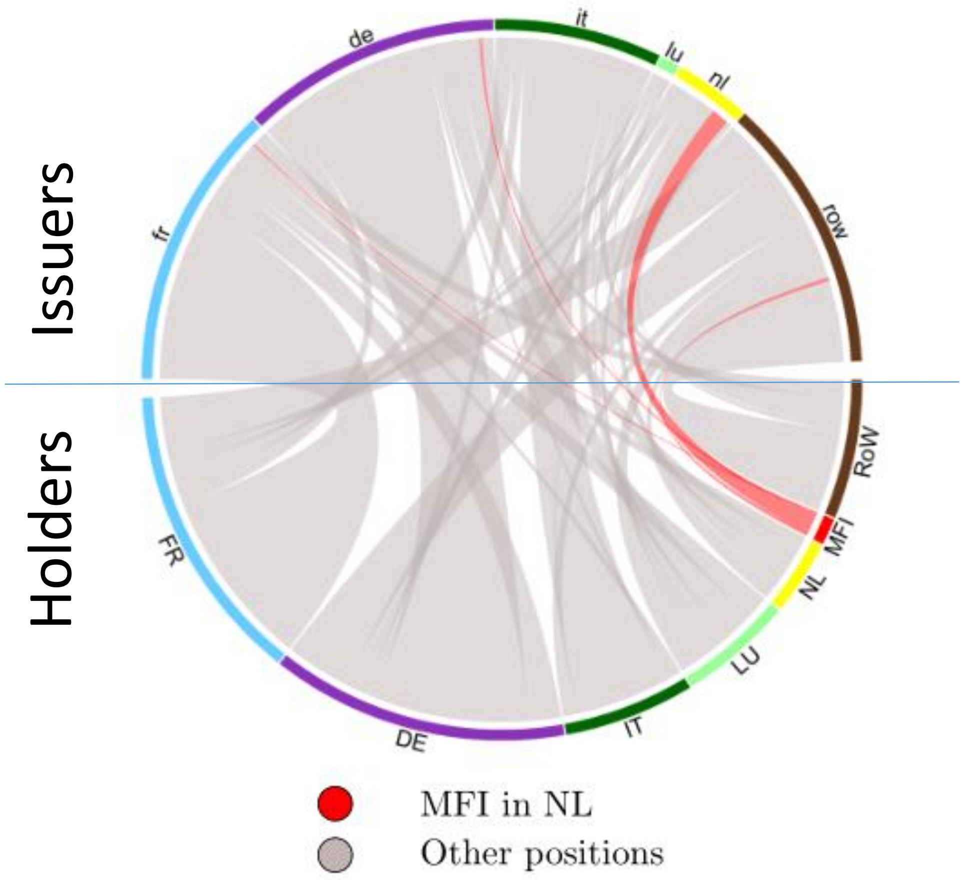
- Data such as SHS has ISIN level exposures on a sector or firm level
- Allows us to see sector-to-sector linkage
- Given security meta data views can extend to focus on:
  - Different securities (eg equity, bonds, …)
  - Maturity
  - Green footprint --> carbon stress test
  - Bond type (eg. CoCo bonds)
- Linkage with country and sector level information
- Macro prudential stress tests

The same granular data can serve multiple goals



- Same data, but now identifying granular sectors
- Allows to analyse the relative risk for different sectors
- Meso stress test

The same granular data can serve multiple goals



- Drilling further into one sector
- Granularity available but confidentiality precludes showing this detail
- Easy linking to prudential information such as balance sheet or risk information
- Here we could do micro prudential stress tests

Considerations

- The need for causal stories
    - We need to open the black box and have actionable stories
- Find your way in the cloud
    - DNB Data Science Hub has put confidential data in the cloud
- Attract new staff and train existing staff
    - DNB Data Science Hub
    - Traineeship “Data and Technology”
    - Learning trajectory: *Become a “Datapreneur”*
- Let data scientists talk to supervisors and policy officers
    - Joint projects in a hub-and-spokes model

WORD DATAPRENEUR!

- Leertraject
- 6 maanden, 1 dag per week
- Los je eigen data-vraag op!

Data-entrepreneurs in de dop gezocht!

Wil jij zelf aan de slag met data om vraagstukken uit de praktijk op te lossen? Geef je dan op voor een nieuw leertraject: Word Datapreneur. Wie weet mag je wel jouw eigen data-oplossing gaan uitwerken... Iman van Lelyveld en Jasmijn van Loo van de Data Science Hub leggen uit.

Wat is ‘Word Datapreneur’?
“Word Datapreneur is een leertraject waarbij DNB’ers een businessprobleem kunnen aandragen dat ze zelf willen aanpakken met behulp van data. Uit de aanmeldingen selecteren we een aantal deelnemers, die vervolgens samen aan de slag gaan om hun eigen en elkaars ideeën te onderzoeken en oplossingen te bouwen. Dit onder begeleiding van collega’s van de Data Science Hub.”

Waarom dit leertraject?
“Met ‘Word Datapreneur’ willen we twee dingen bereiken. Ten eerste werken we aan oplossingen voor bestaande problemen. Ten tweede leer je als deelnemer spelenderwijs coderen, om te gaan met data en samen te werken. Je ontwikkelt je tot data-entrepreneur: iemand die ondernemend bezig is met data, ideeën heeft over hoe data een oplossing kunnen bieden voor beleidsvragen, en die ideeën weet te verkopen. Het geeft handen en voeten aan onze ambitie om digitaal vaardig te worden.”

Wat houdt het programma concreet in?
“We selecteren zes tot tien groepjes die aan hun businessprobleem gaan werken. Zij beginnen half mei en gaan een halfjaar lang één dag per week aan de slag. Het

DeNederlandscheBank
EUROSYSTEEM

Thank you for your attention

Questions?

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Cross-border interbank contagion risk to the Malaysian banking system[1]

Harikumara Sababathy and Lim Sheng Ling,

Central Bank of Malaysia

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Cross-border Interbank Contagion Risk to the Malaysian Banking System[1]

Harikumara Sababathy and Sheng Ling Lim[2]

Abstract

The Malaysian banking system is not immune to potential financial instability induced by greater cross-border interconnectedness, a key feature of the recent Global Financial Crisis. In view of this, a comprehensive and reasonably accurate assessment of cross-border contagion risks is crucial to assess risks to financial stability. In this paper, we utilised network analysis and counterfactual simulation to (i) conduct an in-depth study of the dynamics of Malaysian banks' cross-border and domestic interbank exposures over a period of 5 years and (ii) assess potential losses arising from contagion induced by external shocks. We find that interconnectivity and solvency contagion risk in the Malaysian banking system to external shocks have declined significantly since 2013, despite the rapid growth in cross-border exposures in value terms. Nonetheless, due to their large size and interconnectedness i.e. importance as a financial intermediator, systemically important banks are more susceptible to induce contagion within the network.

Keywords: Contagion, network analysis, interconnectedness, DebtRank, stress test

JEL classification: D85, G01, G17, G21, G28, L14

[1] This paper has been published as part of The South East Asian Central Banks Research and Training Centre (SEACEN) collaborative research project titled "Cross-border Interbank Contagion Risk Analysis" (SEACEN, May 2019). An additional section has been added since to reflect the extension of our work. The views expressed in this paper are those of the authors and do not necessarily represent those of Bank Negara Malaysia (BNM) or its policies. The authors would like to thank Qaiser Iskandar Anwarudin, Hafiz Abu Samah, Syairah Hanin Shaharudin, and Nik Ahmad Rusydan bin Nik Hafizi for their invaluable feedback.

[2] The authors can be contacted via the following email addresses: harikumara@bnm.gov.my and sheng@bnm.gov.my

1. Introduction

In recent times, more so after the Global Financial Crisis (GFC), regulatory authorities worldwide have taken a keen interest in being better able to assess risks of contagion events from occurring. This is because effects of financial contagion was one of the key features of the GFC as localised distress in specific segments of financial markets rapidly morphed into a crisis of global proportions. The OECD (2012) estimates that contagion shocks induced via bank balance-sheet interconnectedness, which had been an important driver of financial crises, reached unprecedented heights during the GFC. True enough, the extent and nature of international banking integration which contributed to an unprecedented transmission of financial instability, was one factor that made the recent financial crisis so deep and widespread.

While most global banks held back cross-border banking activity in the aftermath of GFC, banks in Asia Pacific stepped up, intensifying their cross-border activities particularly within the region, with financial centres such as Hong Kong SAR and Singapore playing important roles (Remolona and Shim, 2015). Notwithstanding the benefits, Remolona and Shim (2015) go on to highlight three potential source of risks to financial stability caused by greater banking integration in Asia: (i) growing systemic importance[3] of foreign banks in host jurisdictions, both as common and concentrated lenders within the region; (ii) high reliance on wholesale and derivative markets for foreign currency funding and (iii) shortening maturity of foreign currency borrowing.

There are several factors which makes the Malaysian banking system susceptible to potential financial instability induced by greater cross-border interconnectedness. First, the Malaysian banking system has one of the highest foreign bank participation among most Asian peer jurisdictions, with foreign banks accounting for 19.8% of total domestic banking system assets as at end-2017. Second, Malaysian banks' cross-border exposures, particularly in the form of short-term borrowings, has increased rapidly in recent times in tandem with growth of domestic banking groups in the regional markets and strong domestic presence of foreign banks. Since 2008, Malaysian banks' external debt almost tripled to RM316.8 billion or 23.4% of GDP as at end-2017. Third, Malaysian banking system is both concentrated and large compared to the domestic economy; with top five largest banking groups by asset size, which also has significant cross-border presence, accounting for 70% of total banking system assets and 231% of Malaysian GDP as at end-2017.

[3] Refers to banks whose distress or failure may result in severe negative spillover impact to the financial system and real economy. At the global stage, the Financial Stability Board and Basel Committee of Banking Supervision assess and identify banks that are deemed to be systemically important on an annual basis, referred to as global systemically important banks (G-SIB). Similarly, national authorities identify banks that are systemically important in their respective jurisdiction, commonly referred to as domestic systemically important banks (D-SIB).

2. Literature review

There are two main strands of literature related to our work. The first discusses the use of network analysis to assess contagion risk while the second deals with the use of counterfactual simulation to quantify the impact of contagion risk.

2.1 Network Analysis

Network analysis has been increasingly recognised as a powerful surveillance tool for modelling the interconnectedness of the banking system network, although not widely employed due to data limitations. Hattori and Suda (2007) explored the 'core-periphery' network of cross-border bank exposures for 215 countries using the Bank for International Settlement (BIS) locational banking statistics data for the period 1985-2006. They found that the network has become more tightly connected with higher average degree, higher clustering coefficient and shorter average path length over time. In addition, network features remain largely undisrupted by any disturbances or crises in the international financial markets. Although systemic risk build-up is inevitable, financial markets are allocating capital and risk more efficiently.

Using similar dataset, Minoiu and Reyes (2011) analysed interconnectivity of the global banking network for 184 countries over 1984-2009. They enhanced existing literatures on network study through the use of cross-border flows data to reflect liquidity conditions and applied the weighted degree approach to capture the heterogeneity in cross-border exposures. The authors posited that network density tends to expand and contract in line with the cycle of capital flows. Furthermore, country centrality decreases during and after banking and sovereign debt crises, with the GFC standing out as an unusually large disruption to the global banking network, contradictory to findings from Hattori and Suda (2007).

Utilising data on bilateral cross-border exposures between 1980 to 2005 for 18 advanced and emerging economies, Kubelec and Sa (2010) found increasing financial interconnectivity over time. The global financial network is clustered among few key nodes with large financial links and lower average path length over time, focusing around the United States and United Kingdom as its central hubs. Due to its robust yet fragile network structure, disturbances to the key hubs would transmit shocks rapidly and widely throughout the network. Comparison with the trade network shows similar increase in interconnectivity over time. Nonetheless, the trade network exhibits strong intracontinental links, concentrating around three clusters, the European cluster (centred on Germany), American cluster (centred on United States) and Asian cluster (centred on China).

2.2 Counterfactual Simulations

While network analysis has aided our understanding on risks posed by greater interconnectedness, actual occurrences of default cascade events (referring to instances of a bank's default impacting or inducing default of other banks via direct balance sheet linkages or indirect linkages such as precipitation of bank run following loss of confidence) remained rare even during the height of GFC, largely due to widespread regulatory intervention. This has led to the lack of reliable empirical evidence, which allows for reasonably accurate estimation of such risks and impact to the financial system and economy upon materialisation of such events. Against this backdrop, counterfactual simulations have emerged as an important approach to assess the likelihood of contagion occurring (Upper, 2011; Elsinger, Lehar and Summer, 2012). Such simulations are increasingly being used by regulators either on a stand-alone basis to assess resilience of a particular financial system or as part of broader macroeconomic stress testing approaches, for example the Bank of England's RAMSI (Burrows et.al., 2012).

Upper (2011), which provides an excellent summary of counterfactual simulation studies of the bank-driven contagion, specifically via the interbank market channel, found two major shortcomings in the reviewed literature: (i) an exaggerated focus involving the idiosyncratic failure of an individual bank in constructing shock scenarios, and (ii) the general neglect of mechanisms that extends beyond direct balance-sheet linkages. Such mechanisms like the amplification of losses by behaviour and asset pricing, which can be significant during crisis periods[4], can lead to misleading conclusion that contagion risk is small. Therefore, it was not surprising that the early literature on counterfactual simulations using interbank exposures established the following two premises: (i) contagion of insolvency due to interbank exposures are rare and (ii) it is difficult to create realistic scenarios that will induce significant amount of contagion (Summer, 2013). Much recent literature however has emphasised the importance of considering other contagion channels in constructing counterfactual simulations. Glasserman and Young (2015) highlights the importance of mechanisms which goes beyond simple spillover effects (referring to default cascade approaches) to magnify shocks such as (i) bankruptcy costs and (ii) valuation losses resulting from deterioration of counterparty creditworthiness or a loss of confidence. In particular, they found that loss of confidence can lead to widespread losses of value – with their analysis suggesting this channel of contagion is likely to be more important than simple spillover effects.

Put differently, a possible explanation for the general neglect of other forms of contagion mechanisms could have been the reliance on default cascades to propagate distress within networks i.e. distress within the network is only induced upon the failure of a bank in response to either a common shock or idiosyncratic events, a rather common feature across early counterfactual simulation (refer to Upper, 2011; Elsinger, Lehar and Summer, 2012; Glasserman and Young, 2015). Beyond being an unrealistic assumption considering how recent financial crises events have proven[5], default cascade approach turned out to be problematic as it was difficult to induce contagion having bank defaults as a precondition, limiting the usefulness of such simulation studies. For example, given banks are typically subjected to single counterparty exposure limits which typically falls well below banks' total capital, a simulation study with shock scenario involving a single defaulting bank would neither be sufficient to induce other banks' default nor induce widespread contagion. To induce meaningful contagion in such studies, one has to create a shock scenario with simultaneous instances of bank defaults within the financial system, which are both rare and again, unrealistic in actuality.

In this regard, development of DebtRank methodology motivated by work on distance to default, which enabled propagation of distress within network without

4 BCBS (2018) estimates that mark-to-market losses from credit quality deterioration exceeded losses incurred from outright defaults during the GFC.

5 For instance, the onset of Global Financial Crisis and ensued contagion were largely driven by sudden and pervasive averseness following a general loss in confidence as opposed to widespread defaults.

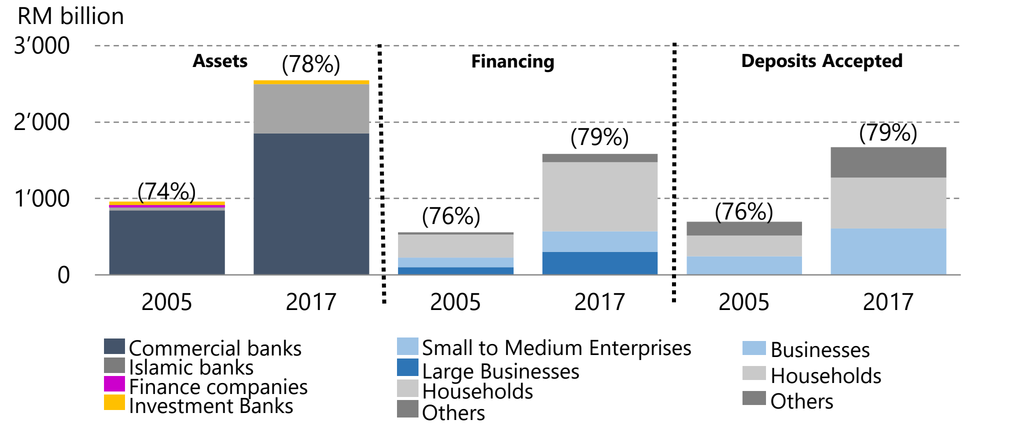
assuming defaults (Battiston et. al., 2012) gained prominence as a method to estimate the impact of shocks within financial networks (Aoyama et. al., 2013; Tabak et. al. 2013; Bardoscia et. al., 2015; Fink et. al., 2016; Bardoscia et. al., 2017). Of significance, DebtRank methodology using relative loss of bank equity within a network as a measure of distress, enabled the propagation of distress induced by valuation losses within networks following deterioration of counterparty creditworthiness or a loss of confidence to be incorporated in simulation studies. Consistent with this, the impact of shocks measured using DebtRank is typically larger than the traditional default cascade approach given the ability to induce further losses within the network. Battiston et. al. (2015), which incorporated DebtRank to carry out a stress-test exercise on interbank exposures of 183 listed European banks over the years 2008-2013 found that the second-round and third-round effects of distress dominate first-round effects, further underscoring the importance of taking into account other contagion channels/mechanism in counterfactual simulation. The original DebtRank assumed that (distress and thus) losses are propagated linearly between connected bank. This assumption was subsequently relaxed in later studies (Bardoscia et. al., 2016; Bardoscia et. al., 2017) to introduce other non-linear propagation functions.

3. Malaysian Banking System

The Malaysian banking sector is well-developed and diversified, with a wide range of domestic and foreign financial institutions. At the end-2017, the banking system comprises 27 commercial banks (8 domestic and 19 foreign), 16 Islamic banks (10 domestic and 6 foreign) and 11 investment banks, with total assets worth approximately RM2.5 trillion, or close to 200% of GDP (Graph 1). Commercial banks are the largest providers of funds, performing retail banking services, including acceptance of deposits, supplying loans and providing payment and remittance services. Islamic banks conduct a similar range of banking activities based on Shariah principles. Meanwhile, investment banks are responsible as financial intermediary for capital market activities such as dealing in securities, raising capital and security underwriting besides providing corporate advisory services. In terms of financing, the banking system has about RM1.6 trillion in outstanding loans, equivalent to over 117% of GDP as at end-2017.

Composition of Assets, Financing and Deposits Accepted for the Banking System

Graph 1



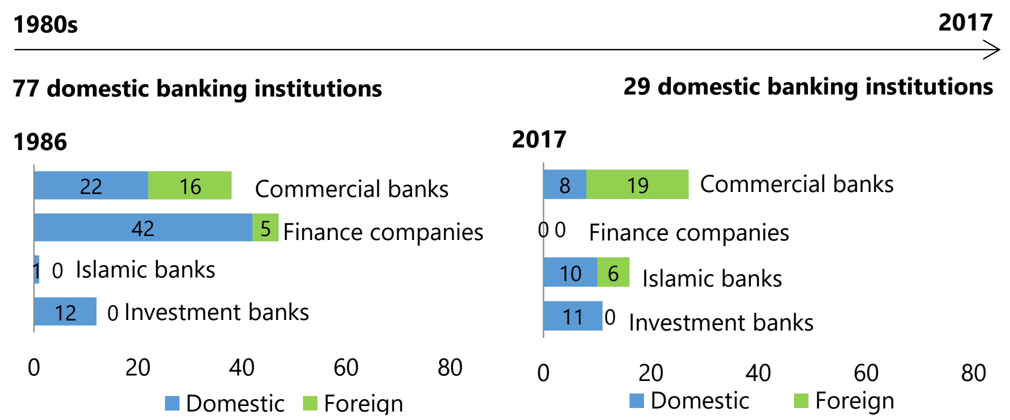
Note: Figures in parentheses refer to market share of domestic banks relative to banking system

Source: Bank Negara Malaysia

Considerable efforts have been undertaken to strengthen the resilience of the banking system over the last decade, particularly in the aftermath of the Asian Financial Crisis. Realising the needs for stronger and larger domestic financial institutions, an industry-wide restructuring and consolidation programme was introduced by BNM in 1999. This has successfully reduced fragmentation and transformed the banking industry to better equip them with enhanced capacity to serve the domestic economy. The 77 domestic banking institutions that existed in the 1980s have merged to form 8 main domestic banking groups (Graph 2) under the consolidation programme. As highlighted in the introduction, the top five largest domestic commercial banks in Malaysia now account for 70% of total of total banking system assets. 5 out of 8 domestic banking groups have also expanded regionally and emerged as key players within the ASEAN region.

Consolidation of the Financial Sector

Graph 2



Source: Bank Negara Malaysia

Although banks in advanced economies faltered during the GFC – with some ending in bankruptcy – the Malaysian banking system remained resilient with sound profitability levels and ample capital buffers to absorb shocks in the event of future stressed events. As at end-2017, the capital ratios of all banks were well above the regulatory minima. More than 75% of capital are in the form of highest quality loss-absorbing instruments, which includes equity, retained earnings and reserves, allowing banks to withstand macroeconomic and financial shocks without adversely affecting financial intermediation. A series of macroprudential measures implemented since 2010 and strengthened risk management practices of banks also further mitigated potential risks by reducing banks' exposures to over-leveraged households (Bank Negara Malaysia, 2017), thus improving the asset quality of banks to a low of 1.1%. Liquidity Coverage Ratio (LCR) requirements have been phased in since June 2015 for the banking institution as part of the Basel III reform package, which required all banks to maintain sufficient high-quality liquid assets at a minimum of 100%[6] of total projected net cash outflows for the next 30 days. Banks have all transitioned smoothly to this requirement.

Key Financial Indicators of Banks — Table 1

Banking System	2000	2017
Capitalisation		
Risk-Weighted Capital Ratio/Total Capital Ratio (%)	12.4	17.8
Asset Quality		
Net NPL Ratio (%)	9.9	1.1
Profitability		
ROA (%)	1.5	1.5
ROE (%)	19.5	13.1
Liquidity		
Liquidity Coverage Ratio (%)	-	134.9

Note: see the Appendix for detailed formulae.

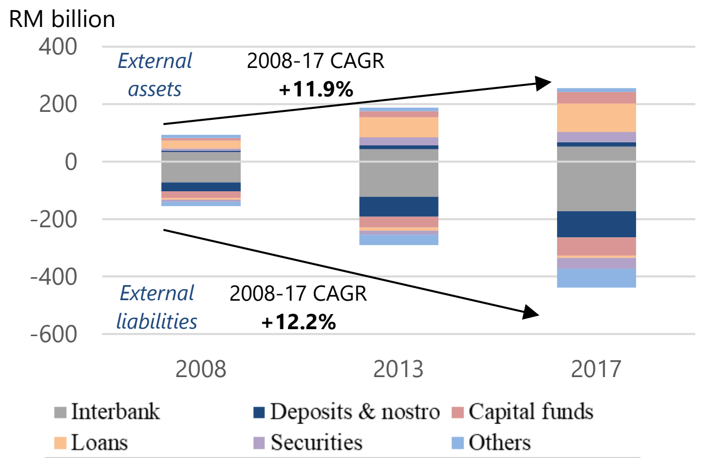
The nature and extent of financial linkages of Malaysian banking system has evolved and increased steadily since the GFC. While total external exposures[7] of Malaysian banks remain low, on aggregate accounting for only 10.2% and 17.2% of the total assets and funding of the banking system in 2017 respectively, banks' claims on, and liabilities obligations to, non-resident counterparties grew at a compounded annual growth rate (CAGR) of 11.9% and 12.2% since 2008 (Graph 3). The rapid expansion Malaysian banks' external exposures is driven by the following: (i) the sizeable presence of foreign banks in Malaysia (Graph 4), including operations in

[6] The LCR requirement was phased-in from June 2015, with an initial transitional requirement of 60%. Full compliance of 100% requirement is in effect starting 1 January 2019.

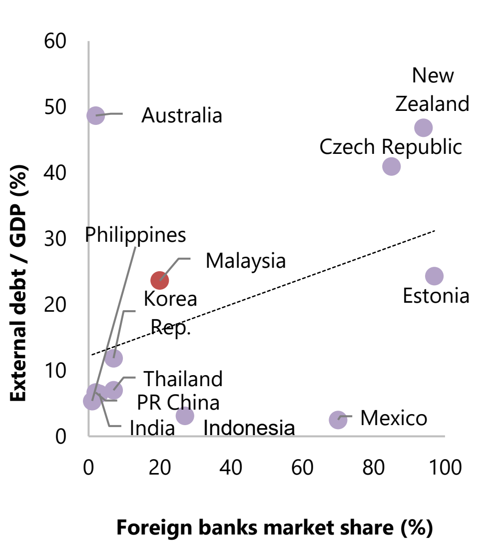
[7] We shall be using the term 'external exposures' and 'cross-border exposures' interchangeably in this paper. Both refer to banks' exposures vis-à-vis non-resident counterparties.

Labuan International Banking and Financial Centre (LIBFC)[8] and (ii) regionalisation of domestic banks, particularly in ASEAN region (Graph 5). In line with this, bulk of the external exposures are in the form of (i) intra-group transactions between banks and related offices[9] abroad in the form of interbank placements and borrowings and capital funds, (ii) deposits accepted from non-residents and (iii) loans extended to non-resident counterparties.

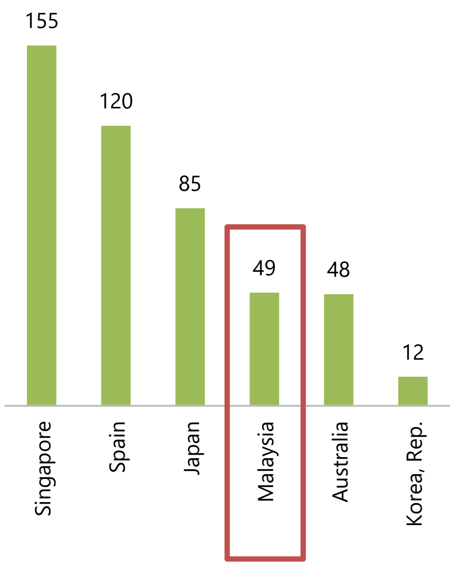
Banking System External Exposures — Graph 3



External Debt of Banks/GDP vs Foreign Banks Market Share in Banking System — Graph 4



Total Foreign Claims of Domestic Banks (% of GDP) — Graph 5



8 LIBFC, established in 1990 is an offshore financial centre, which offers tax incentives for financial and capital market activities. In lieu of this, LIBFC banks' role as a booking centre is prominent, as onshore banks, primarily domestic banking groups, conduct their FCY intermediation activities via their Labuan offices. We shall be using the terminology LIBFC and Labuan interchangeably in this paper.

9 For foreign banks, related offices refer to overseas parent and regional offices. For domestic banks, related offices refer to overseas subsidiaries and branch operations.

Note: Foreign bank market share data refers to proportion of total banking sector assets held by foreign banks Source: Global Financial Development Database, Bank Negara Malaysia, Haver Analytics, Bloomberg, IMF World Economic Outlook	Note: Refers to total foreign claims of domestic banks in all currencies. Data as at end-3Q 2018. Source: BIS International Banking Statistics, Bank Negara Malaysia, Haver Analytics, Bloomberg, IMF World Economic Outlook

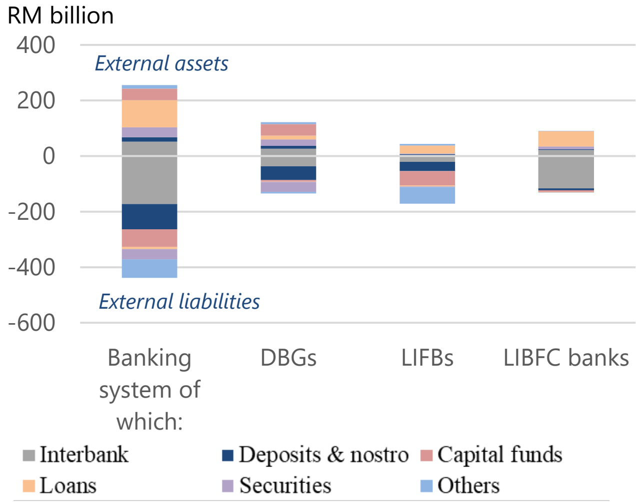
Of Malaysian banks' external exposures in 2017 (Graph 6):

(i) 48% and 31% of external assets and liabilities were attributed to domestic banking groups (DBGs), with significant regional presence. Bulk of the exposures relates to centralised liquidity management practices[10] of DBGs. For DBGs, liquidity is typically managed at the head office to manage maturity and currency mismatches across the group operations. Excess liquidity from related offices along with FCY funding raised in international capital markets as part of the broader funding diversification strategy are redistributed back to related offices.

(ii) 35% and 30% of external assets and liabilities were attributed to LIBFC banks. External exposures of LIBFC banks predominantly reflect placements by related entities to fund foreign currency (FCY) intermediation activities. Such reliance on funding from related offices is expected for offshore banks, which typically operates as a branch and has limited alternative funding sources.

(iii) 17% and 39% of external assets and liabilities were attributed to locally-incorporated foreign banks (LIFBs). LIFBs typically leverage on the stronger credit rating of internationally-active parent banks to source cheaper FCY funding. These are primarily channelled towards FCY lending in the domestic interbank market, short-term investments and to manage maturity mismatches in their FCY loan portfolio.

Breakdown of Banking System External Exposures in 2017 by Banks Graph 6

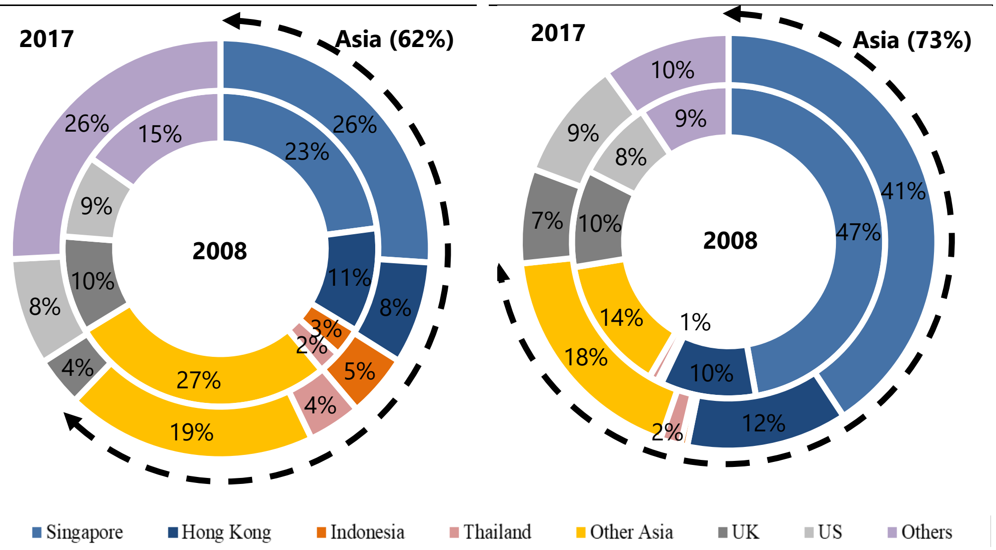


Source: Bank Negara Malaysia

[10] For further information on centralised liquidity management practices of Malaysian banks, refer to Chapter 1 of Financial Stability and Payment Systems Report 2017.

Consequently, from a geographical standpoint, Malaysian banks' external exposures are with counterparties situated in (i) regional financial centres; (ii) countries where Malaysian banks have overseas operations; and (iii) countries with internationally active banks, particularly those with locally-incorporated subsidiaries in Malaysia, with significant concentration observed to Asian region (62% and 73% of cross-border assets and liabilities in 2017) (Graphs 7 and 8).

External assets by location Graph 7 | External liabilities by location Graph 8



Source: Bank Negara Malaysia

Notwithstanding the low risk profile of banks' external exposures (refer to Bank Negara Malaysia's Financial Stability Review, 2018), Malaysian banks are more exposed to adverse market and geopolitical developments across different economies, particularly within Asia. This being as a natural consequence of the rapid growth and profile of banks' cross-border exposures. Moderate spillover effects were felt in the domestic market during periods of tightened US dollar conditions in 2008, as evident from the increased spreads of onshore USD/RM cross-currency swap rates and the widening of short-term US dollar liquidity mismatch positions of banks in Malaysia (Bank Negara Malaysia, 2013). In addition, with sizeable participation of foreign investors, domestic financial markets are also increasingly susceptible to volatile two-way portfolio flows induced by such regional and global developments. This has added to further concerns on banks' external debt[11], which are predominantly in the form of short-term instruments such as deposits and interbank transactions and have evoked concerns of banks also being susceptible to sudden withdrawal shocks.

[11] External debt is a subset of banks' external liabilities, referring to all external liabilities that require payment of principal and/or interest. This is in line with definition of International Monetary Fund (IMF), et al. (2013). In addition to external debt, banks' external liabilities comprise (i) custodial securities held on behalf of non-resident (NR) clients, (ii) capital funds maintained by NR in resident banks, primarily for LIFBs and (iii) financial derivative balances.

4. Data

This paper employs domestic and cross-border interbank exposures data sourced from two separate quarterly reporting submissions: (i) External Assets and Liabilities (EAL) for external cross-border exposures, which can be found in the Bank for International Settlements (BIS) locational banking statistics database for Malaysia, and (ii) Interbank Exposure for information on banks' bilateral exposures in the domestic interbank market. As part of the Interbank Exposure submission, banks are required to submit information on their interbank claims and placements by currencies and counterparties (domestic and external), including placements and borrowings with BNM. Information on bilateral exposures (e.g. counterparty name) to cross-border financial institutions are not readily available for periods before 2017. As such, when looking at cross-border interbank exposures, we use aggregate country level exposures of each bank.

In this paper, we have identified 20 out of 54 banks, comprising domestic and foreign banks, as systemically important to facilitate our analyses. Such banks shall be referred to as systemically important banks (SIBs) hereafter. SIBs are selected based on a combination of balance sheet indicators namely total assets, total deposits and total loans outstanding.

4.1 Network Analysis

There is a total of 54 onshore banks and 24 non-resident countries (including Labuan[12]) in the dataset, forming a network with 78 *nodes*. The network is modelled using interbank lending data and the flow of exposures from a lender to a borrower is considered as a *link*. If credit exposures from one node to another is recorded as 'non-zero', a link is formed between the two nodes. Data was collected on quarterly basis from 2013 to 2017. To ensure consistency in the dataset, the same 54 banks are retained throughout the sample period[13]. Similarly, 21 external counterparty countries have been pre-identified based on their exposures, while the remaining countries clustered accordingly to 3 different regions[14]. These nodes form a directed network with a core-periphery structure, with *core* representing banks who are net lenders while *periphery* comprising the non-resident countries who are net borrowers only. Nevertheless, the core-core network structure (comprising both lenders and borrowers in the domestic interbank market) is also analysed.

Following the footsteps of Minoiu and Reyes (2011), we applied the same treatment on the bilateral interbank exposures data by retaining only positive flows in this study (corresponding to net placements or net assets), while negative flows (net liabilities) are replaced with zeros. The total claims and liabilities are not differentiated according to currency (domestic currency (i.e. Ringgit Malaysia) vs

12 For the purpose of the network analysis, Labuan banks are deemed as non-resident counterparties.

13 New (old) banks' entrants (exit) into the domestic banking system are removed from the dataset. Thus, this study does not consider the possibility of new links, except between the nodes fixed in the dataset.

14 Refer to Table A1 in appendix 1 for full list of countries nodes.

foreign currency) but aggregated as a whole. The resulting matrix is combined with the external cross-border exposures data for each period.

4.2 Counterfactual Simulation Model

For purposes of the counterfactual simulation model, we use annual data of 54 banks between 2013 and 2017. By using actual interbank exposures data, we partly overcome limitations typically associated with use of estimation techniques[15] to construct bilateral interbank exposures. Other balance sheet information such as regulatory capital, placements with BNM and holdings of liquid securities were sourced from monthly reporting submissions by banks. In contrast to Bardoscia et. al. (2015), we use banks' stock of CET1[16] capital in lieu of total equity.

5. Methodology

This section describes the statistical measures used to explain the network topological characteristics and the counterfactual simulation model used to assess contagion risk.

5.1 Network indicators

5.1.1 Connectivity

Network connectivity is the most commonly used topological measure and is defined as the likelihood of connection between nodes. A non-zero exposure between one node and another node, is known as a link. The total number of nodes and links in a network are identified as n and m, respectively. Since the network is a directed network[17], the probability, p that two nodes share a link is defined as

$$p = \frac{2m}{n\,(n-1)} \tag{1}$$

As the number of nodes in the dataset are fixed, interconnectedness of banks and their respective counterparties is solely influenced by the number of links – the more the number of links, the higher the connectivity.

5.1.2 Clustering Coefficient

Clustering coefficient of a node i is calculated as the ratio of the actual number of directed links between the neighbours of node i ($m_{nn,i}$) over the number of potential links between them. When the total number of neighbours of node i is m_i and the

[15] Referring to techniques utilised to estimate bilateral interbank exposures in absence of actual data. Such techniques tends to distribute aggregate exposures as evenly as possible across institutions given set constraints. A general consensus in the literature is that use of such estimation typically leads to underestimation of contagion effects (Summer, 2013).

[16] CET1 capital is a type of regulatory capital maintained by banks, referring to the highest quality and most loss-absorbing capital instrument under the Basel III capital framework.

[17] For simplification, only directed links (where the edges point in a single direction) are analysed in this paper. Bi-directional links can be considered in future research work.

number of potential links are $(m_i (m_i - 1))$, the likelihood of connection between two neighbours that are directly linked to a node in a directed network is

$$C_i = \frac{2m_{nn,i}}{m_i(m_i - 1)} \tag{2}$$

The neighbours of node *i* are defined as nodes that are directly linked to node *i*. For example, in a directed network of net exposures for the core-periphery structure in Figure A1 (refer to Appendix), if both neighbours, nodes *k* and *j* have a link with node *i*, the likelihood that nodes *k* and *j* are interlinked is known as the clustering coefficient. The clustering coefficient of a network, C can be calculated as the average of the clustering coefficient of each node in the network. Thus, the likelihood of the network to cluster is $C = 1/n \sum C_i$. The higher the coefficient, the more tightly-knit the nodes are together, indicating a 'small-world effect'.

5.1.3 Betweenness Centrality

Betweenness centrality of a node can be defined as the ratio of the total number of shortest paths through node i $(\sigma_{kj}(i))$ to total number of shortest paths from node *k* to j (σ_{kj}). The measure used in this paper has been normalised for comparison purposes between different periods.

$$B_i = \sum_{i \neq k \neq j} \frac{\sigma_{kj}(i)}{\sigma_{kj}} \tag{3}$$

Nodes with high betweenness centrality are often important financial intermediaries as they act as a 'bridge' between nodes in a network. In the network example (Figure A2), bank *G* is recorded to have the highest betweenness centrality between all other banks as it has the most number of times a transaction goes through the node. Being the key financial intermediator, a funding shock to bank *G* would thus be transmitted towards all other nodes connected to it and restricts liquidity for bank *H* and *J*.

5.1.4 Average Degree (Outdegree and Indegree)

The degree of a node computes the number of links for each node. Since the dataset is composed of a directed network, the number of outgoing links from a node, that is the number of counterparties a bank lends to (its debtors), is known as the outdegree (Figure A3). On the other hand, the number of counterparties a bank borrows from (its creditors), also known as the number of incoming links, is referred to as the indegree. The average degree of the network is therefore the ratio of the total number of links, *m* and the total number of nodes, *n*.

Moreover, node strength refers to the value of net interbank exposures originating or terminating at a given node. An increase in the amount of interbank lending by a creditor to its counterparties shows greater node outstrength. Although the number of outward links signals the importance of the bank as a credit provider in the system, the amount of lending should also be emphasised when determining the riskiness of the entity as a credit provider. To put it simply, a bank (A) that lends to a single entity with an amount of RM100 million is more risky than a bank (B) that lends to 5 different counterparties with an amount of RM10 million each. This paper adopted the weighted network approach proposed by Opsahl et. al. (2010) by taking into account the total number of counterparties a bank lends to (outdegree) and their weighted size of exposures (outstrength) using a tuning parameter, α. By assigning α=0.5, both the number of outdegree and weight of the links are given equal relative importance. This centrality measure enables us to more accurately assess the contagion risk of banks with high outdegree relative to their size of interbank placements with their respective counterparties.

5.2 Counterfactual Simulation Model

Bank-driven contagion can take many channels (see Upper, 2011) but this paper focuses on the interbank market, specifically asset-side shocks due to lending and borrowing exposures to financial institutions in the cross-border interbank market (direct effects) and the spillover impacts onto the domestic interbank market via the confidence channel (indirect effects). To assess the resilience of Malaysian banks to susceptibility to shocks induced via the cross-border and domestic interbank markets, a counterfactual simulation adapting the DebtRank methodology developed by Battiston et. al. (2012) shall be utilised. In particular, the DebtRank methodology shall be used to extend the simulation analysis presented in the 2013 Financial Stability and Payment Systems Report (Bank Negara Malaysia, 2013) by attempting to estimate, quantify and track the spillover impact of shocks incurred on banks' cross-border interbank exposures onto the Malaysian interbank system. Results of this extension is explained in Section 8.

5.2.1 Model description

In this section, we introduce and describe the key features of our counterfactual simulation model which had been calibrated by taking into account some of the key insights and shortcomings identified in similar studies highlighted earlier. Detailed information about the process and the main variables of interest can be found in Appendix 2.

Our model envisages a financial system of n banks as a network. Each bank either lends to other banks within the network i.e. domestic interbank lending (A_{dib}), or engage in any of the following (i) cross-border interbank lending to non-resident financial institutions (A_{ea}), (ii) placements with central bank (A_{cb}), (iii) holding of liquid securities (A_{liqs}) and/or (iv) holding of other assets (A_{oth}). Bank assets are assumed to be marked-to-market whereas liabilities are carried at their face value. The rationale being the value of banks' obligations do not change simply because they might not be able to meet the obligations. At any given time, total equity value, E_t of a bank refers to its total assets, A_t less its total liabilities, L.

$$A_t^{dib} + A_t^{ea} + A_t^{cb} + A_t^{liqs} + A_t^{oth} = L + E_t \tag{5}$$

The model is centred on measuring relative equity loss of each bank or the system as a whole at any given time. Given total liabilities are assumed to be carried at face value, changes in value of total assets would result in corresponding change in the value of banks' total equity. The relative or percentage loss in total equity, denoted by h_i is taken as a proxy for distress for any given bank, also referred to as vulnerability of the bank. When $h_i = 0$, a bank is deemed to be un-distressed. All values between 0 to 1 suggests a bank is under distress.

$$\text{Relative loss in equity, } h_{i,t} = \min\left\{1, \frac{E_{i,t=0} - E_{i,t}}{E_{i,t=0}}\right\}, \text{ t} = 0,1,2\ldots \tag{6}$$

At the system level, aggregate vulnerability, H_t refers to the weighted average of each bank's vulnerability, with the weights given by its relative initial equity.

$$\text{System equity loss, } H_t = \sum_i w_i \, h_{it} \tag{7}$$

$$\text{weights, } w_i = \frac{E_{i,\, t=0}}{\sum_j E_{j,\, t=0}} \tag{8}$$

5.2.2 Shock Scenarios

Two types of shock scenarios were used for the simulation. The first scenario assumes banks experiencing a common shock on their cross-border interbank exposures i.e. interbank lending and borrowing exposures to non-resident financial institutions. Specifically, all banks would experience simultaneous credit and funding shocks at time, $t=1$ (also called the first round for brevity); (i) an effective[18] negative credit shock, r on cross-border interbank lending exposures and (ii) a reversal shock, f on cross-border interbank borrowing exposures. Unable to replace the lost funding in the short-run, banks are forced to deleverage, first withdrawing their excess liquidity placements with the central bank, A_{cb} and then liquidating a portion, z of their holdings of liquid securities, A_{liqs} if there are insufficient placements with the central bank. Malaysian banks have high degree of common asset exposures with Malaysian sovereign debt securities forming 40.7% of the banking system's Basel III Liquidity Coverage Ratio High Quality Liquid Assets in 2017. As such, a haircut, x is imposed on all remaining $(1-z)$ stock of liquid securities to replicate the potential price impact of simultaneous asset liquidation to replace lost interbank funding and adverse market movements that may be observed during periods of heavy portfolio outflows that could accompany the interbank funding reversal. This is in part cognisant of the concentration risk posed by emerging markets' limited access to other forms of high quality liquid assets beyond sovereign debt, particularly following the implementation of Basel liquidity frameworks (Basel Committee on Banking Supervision, 2014). Notwithstanding any asset-specific qualities and market dynamics which may offset potential valuation losses during times of rapid deleveraging, we deliberately chose to adopt a conservative approach and impose a haircut on all of banks' liquid assets. The magnitude of the haircut is calibrated based on the historical worst possible adverse yield movement observed on Malaysian government debt.

In the first round, banks will immediately incur the losses following the shocks and there is a corresponding reduction in equity level. The losses in equity at this point is referred to as first order impact. Following this, the likelihood of a distressed bank (referring to a bank suffering from relative loss in equity) repaying its obligations on the domestic interbank market becomes lower as its creditworthiness deteriorates. Here the DebtRank algorithm is adapted to trigger contagion effects in the domestic interbank market whereby banks that lent to distressed banks will incorporate

[18] The effective credit shock on banks' cross-border exposures refers to product of probability of default (PD) and loss-given-default (LGD). For simplicity, the current study only varies the effective credit shock in conducting sensitivity analysis.

information about the reduced creditworthiness into the valuation of their interbank lending exposure thereby incurring losses. We refer to the cumulative loss of equity at this point as second order impact. The value of the interbank asset will lie between its face value (repaid in full) and zero (assuming no recovery in the event of default). In our study, the losses incurred in the second round by creditor banks is proportional[19] to relative loss of equity incurred by debtor banks in the first round. Should a bank have multiple interbank lending exposures to distressed banks in the domestic interbank market, overall losses incurred during the second round refer to total losses incurred on all its domestic interbank lending exposures.

The first scenario is repeated with various values of credit, funding and market shocks to assess sensitivity of impact to set parameters. The second scenario, meanwhile, subjects each bank to similar shocks (as in the first scenario) but individually. This step is then repeated for each bank within the sample. This is intended to decompose systemic risk properties of a bank into (i) its impact on the system and other banks in the event of the bank's distress or default and (ii) its vulnerability with respect to external shocks. Impact of bank i refers to the relative equity loss of the system when bank i is shocked. Both scenarios are then repeated for data ending 2013 and 2017, respectively.

6. Results

The results in this section are divided into two parts: (i) the topological characteristics of the domestic and cross-border interbank network and (ii) the results of the counterfactual simulation. In highlighting our findings, where applicable, we shall focus on comparing (i) SIBs versus non-SIBs and (ii) domestic versus locally-incorporated foreign banks for greater clarity.

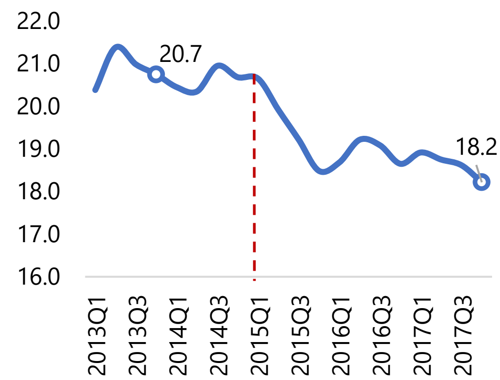
6.1 Network Analysis

Connectivity within the Malaysian banks' interbank exposures has declined over time (Graph 9). Interconnectivity of banks experienced a relatively steep decline in 2015 before reaching 18.2% in 2017, compared to 20.7% in 2013. As the number of nodes in the dataset are fixed, the decline in connectivity is a factor of the total number of links in the network, which dropped to an average of 559 for the year 2017 from 627 in 2013. Further breakdown shows this is largely attributed by lower connections from SIBs, mainly made up by domestic players. The average number of transactions made by these domestic SIBs fell to an average of 281 in 2017 (2013: 349).

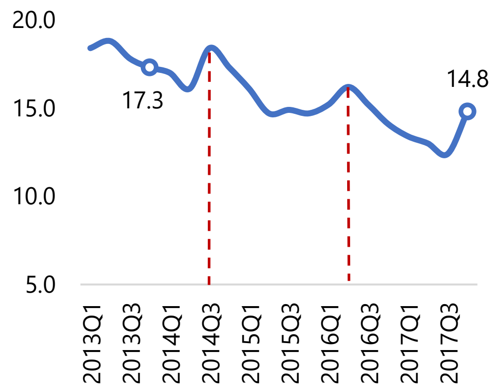
[19] Here we would like to highlight that that subsequent works on DebtRank have advanced refined valuation functions for various asset exposures which can more accurately calibrate the distress propagation mechanism (Bardoscia et. al., 2017). In our view, although the extent of the over- and underestimations of using a linear propagation mechanism cannot be rigorously quantified, a reasonable assumption is that they roughly cancel each other out.

Connectivity (%) Graph 9	Clustering Coefficient Graph 10

Connectivity (%) Graph 9



Clustering Coefficient Graph 10

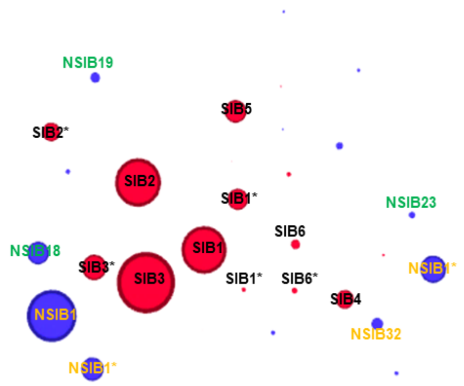


Source: Author's estimation

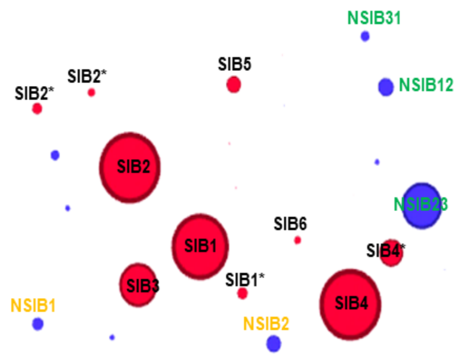
Similarly, the likelihood that two neighbouring counterparties are connected are also on a declining trend, with the clustering coefficient estimated at 14.8% in 2017 (Graph 10). There are generally two bouts of decline observed during the period, mainly in 3Q 2014 and mid-2016. This is largely due to outflows, particularly to Labuan. The results above signals that although lesser links are generated within the interbank market over time, the network is prone to movement of flows, observed by the increase in intensity of cross-border flows.

The change in betweenness centrality from 2013 to 2017 confirms that the role of some key SIBs remained important, recording a betweenness centrality of 0.05 to 0.08 as at 4Q 2017[20]. The four largest parent SIBs remain as key financial intermediaries in the domestic market despite the declining role of their subsidiaries (Figures 1 and 2). Amongst the list of banks with the top ten highest betweenness centrality statistic, two of the banks are LIFBs consistent with the significance of foreign players as a provider of foreign currency liquidity.

1Q 2013 Network: Betweenness centrality Figure 1



4Q 2017 Network: Betweenness centrality Figure 2



Note:

20 Refer to Table A2 (Appendix 3).

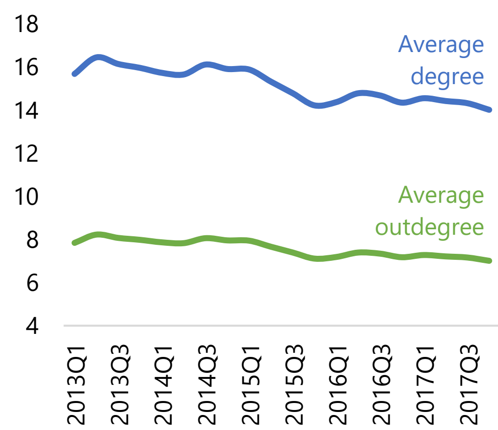
1. Size of the node represents measure of betweenness centrality i.e. the larger the node, the higher the betweenness centrality
2. Red nodes represent SIBs, blue nodes represent non-SIBs
3. Nodes with green labels represent non-SIB LIFBs, nodes with yellow labels represent non-SIB domestic banks
4. *Subsidiaries of SIBs

Source: Author's estimation

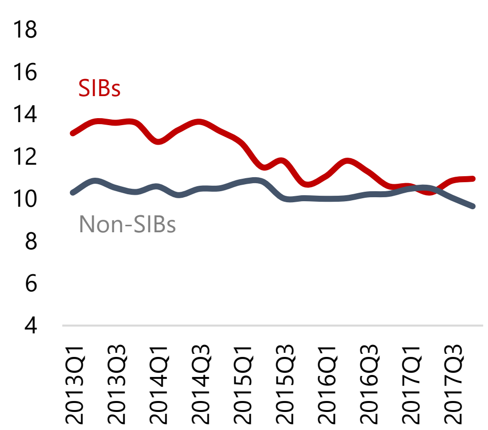
The average number of outgoing links has also been on a steadily declining trend, reflecting the decrease in interconnectivity of banks (Graph 11). Specifically, the fall in the average degree is observed since 3Q 2014. Narrowing on the domestic interbank market only (the core-core network), given that there are 54 banks, the maximum number of potential outgoing links in this study would be 1,431[21]. However, given that there are only 169 directed-out links in 2017, only 12% of the potential links are utilised, deeming the domestic interbank system as an incomplete network. This is not much different from 2013, where there were 216 outgoing links, thus exploiting 15% of the potential core-core network. A look at the core-periphery structure, however, shows a slightly more complete network, with 54% of the potential outgoing links between banks and their respective external counterparty countries being utilised as at end-2017. This is reasonable because the greater regionalisation of domestic banks, deep and liquid financial market coupled with strong economic performance attracted a larger presence of foreign banks.

The decline in average outdegree is more apparent for SIBs, albeit remaining higher than non-SIBs (Graph 12). This is largely attributed to the increase in cross-border flows, where more than half of exposures are to non-residents (Graph 13).

Average Degree and Average Outdegree Graph 11



SIBs and Non-SIBs: Average Outdegree Graph 12



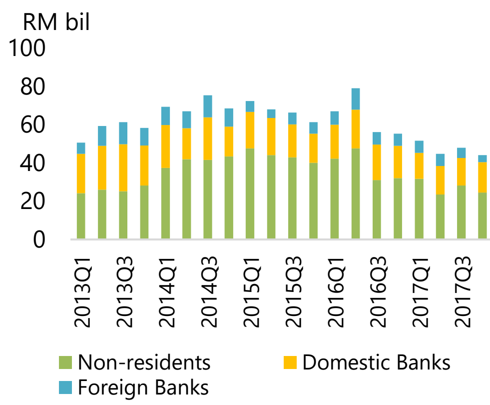
Source: Author's estimation

Further breakdown of net interbank exposures of SIBs to non-residents show that the largest exposure of our onshore banks was to the Asian region, specifically Labuan

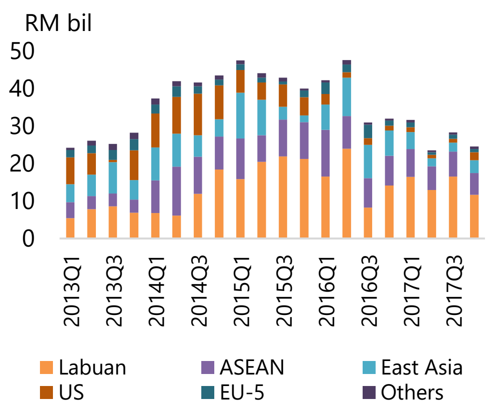
[21] The maximum number of potential outgoing links is as follows = $\frac{54 \times 53}{2}$

(Graph 14). Intragroup funding to Labuan expanded in mid-2014 following the increase in foreign currency lending to corporations to support business activities during the period. The general decrease in cross-border exposures observed in 3Q 2016, however, is partly explained by some DBGs' strategy to reduce reliance of overseas operations on parent funding, limiting external exposures.

1Q 2013 Network: Betweenness centrality — Graph 13



4Q 2017 Network: Betweenness centrality — Graph 14



Source: Bank Negara Malaysia

Table 2 shows the top 10 net exposures of D-SIBs to non-residents by countries. Net interbank placements in Labuan as at 1Q 2013 was recorded at RM5.4 billion. However, this figure ballooned to RM11.7 billion as at end-2017, showing the increasing role of Labuan as a booking centre. Cross-border interbank claims continue to be concentrated within the Asian region, particularly in countries dominated by our domestic banks' regional presence.

SIBs: Top 10 Net Exposures by Countries (1Q 2013 vs 4Q 2017) — Table 2

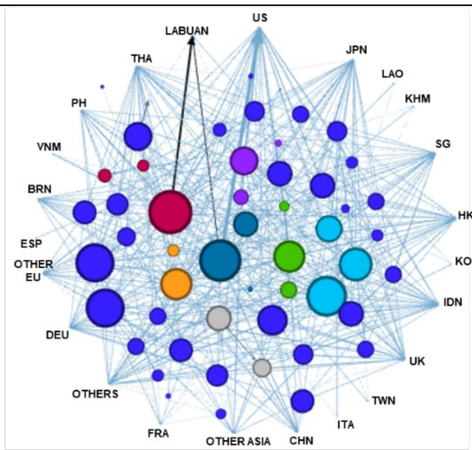
No	1Q 2013		4Q 2017	
	Country	Exposure (RM bil)	Country	Exposure (RM bil)
1	US	7.1	Labuan	11.7
2	Labuan	5.4	Singapore	3.1
3	Hong Kong	2.5	US	2.2
4	Singapore	2.2	China	1.6
5	China	1.2	Philippines	1.1
6	Indonesia	0.9	Hong Kong	0.8
7	France	0.8	Indonesia	0.6
8	UK	0.8	Japan	0.6
9	Vietnam	0.5	UK	0.5
10	Philippines	0.5	Korea	0.5

Source: Bank Negara Malaysia

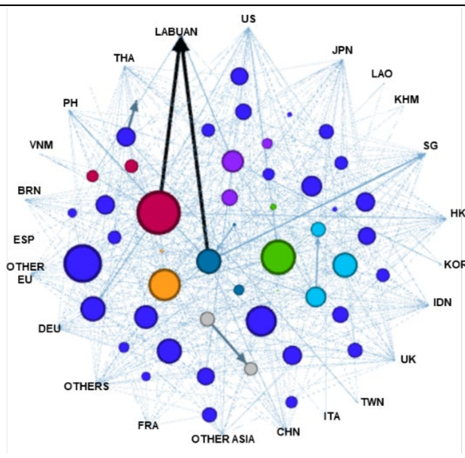
The time evolution of the interbank network is illustrated in Figures 3 and 4. The network is shown at the beginning and end of the sample period (1Q 2013 vs 4Q 2017). There is a clear decline in network connectivity in the domestic and cross-border interbank market over time, suggesting a decline in the network susceptibility to contagion events. The decline in interbank connectivity can be explained by the following: (i) greater cross-border flows to specific non-residents by SIBs and (ii) concentration of interbank activities with related counterparties (i.e. intragroup exposures), a less risky source of exposure.

The LCR requirement, which was first announced in November 2014 could have contributed to the diminishing connectivity within the interbank network. To meet the requirements, banks are required to equip themselves with HQLA which includes placements with central banks and holding of liquid debt securities. In this regard, banks would have been incentivised under the requirement to increase placements with BNM (i.e. lending to BNM in the domestic interbank market) as opposed to lend to another financial institution within the domestic interbank market as the former counts as a HQLA, thus lowering interconnectivity. Following the implementation of LCR requirement in June 2015, we did observe a general decline in interbank lending by banks while placements with BNM[22] rose and remained at elevated levels thereafter (Graph 15). In addition, the 100% run-off rate attached to interbank borrowings, which effectively means banks have to increase their stock of HQLA by an equivalent amount of their borrowings, could have dis-incentivised banks to borrow from the interbank market, thereby also lowering interconnectivity.

1Q 2013:
Network Outdegree — Figure 3



4Q 2017:
Network Outdegree — Figure 4



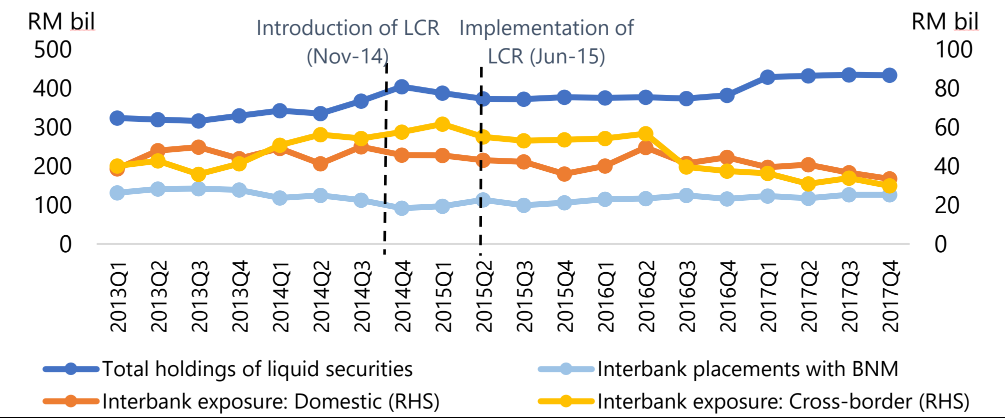
Nodes legend:
SIB 1 SIB 2 SIB 3 SIB 4 SIB 5 SIB 6 SIB 7

Note:
1. Royal blue nodes represent non-SIBs
2. Size of the nodes represent the number of counterparties
3. Direction and thickness of the arrows represent net placement and relative size of exposures

Source: Author's estimation

22 Central bank placements could be argued to be a safer option for banks compared to liquid securities holdings especially during times of volatility in the domestic financial markets with increased two-way portfolio flows, as was the case during 2015.

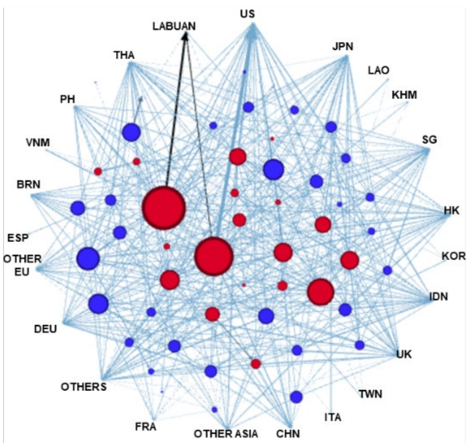
Average Degree and Average Outdegree Graph 15



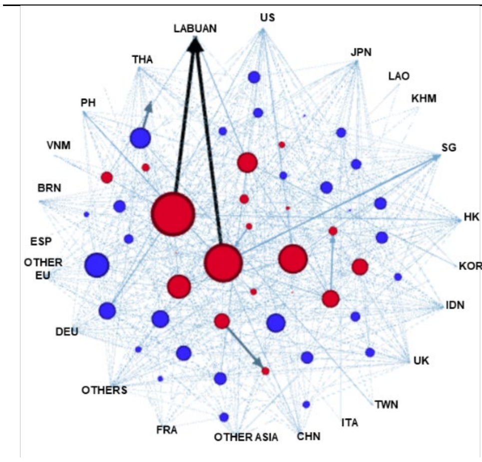
Source: Bank Negara Malaysia

Using a weighted network approach, we find that the risks for some SIBs have decreased slightly over time in line with diminishing connectivity (Figures 5 and 6). It is interesting to note also that contagion risk used to be widely distributed amongst SIBs, but now the risk is more concentrated within the larger SIBs. Nonetheless, SIBs remain relatively more susceptible to induce contagion compared to non-SIBs due to their large size and important role as a credit intermediator.

1Q 2013 Network: Weighted Network Outdegree Figure 5



4Q 2017 Network: Weighted Network Outdegree Figure 6



Note:
1. Red nodes represent SIBs, blue nodes represent non-SIBs
2. Size of the nodes represent the degree of interconnectedness, taking into account the number of counterparties and weighted size of exposure
3. Direction and thickness of the arrows represent net placement and relative size of exposures

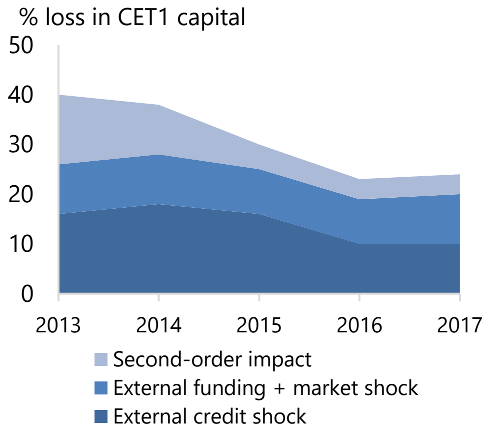
Source: Author's estimation

6.2 Counterfactual Simulation Analysis

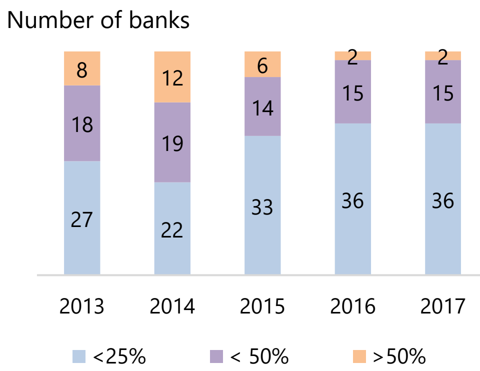
In assessing the results of counterfactual simulation model, we start by exploring the losses incurred at the aggregate system level over the five-year period. Recall that this period also coincides with several key structural changes that took place in the Malaysia banking system: (i) implementation of Basel III capital and liquidity standards and (ii) a significant decline in the aggregate interbank exposures in the domestic interbank market. Graph 16, which offers a deconstruction of the losses according to shock types and rounds the losses were incurred in, shows that aggregate losses due to contagion has decreased significantly from 2013 to 2017. Consistent with findings of lower connectivity in the domestic interbank market, the decline in aggregate losses are driven by lower loss contribution from second round effects. That said, relative losses in equity due to the second round continue to be sizeable, suggesting that disregarding spillover or indirect impact can underestimate the true extent of potential losses. In addition, valuation losses incurred on liquid securities holdings arising from rapid deleveraging to replace lost interbank funding have steadily increased over the years, making up 42% of total aggregate losses at the system level in 2017 as compared to 25% in 2013. While the banking system cross-border interbank funding has grown rather significantly in value terms during the same period, we have maintained the same haircut for liquid securities holdings for simplicity. Given this, the increase in valuation losses is primarily attributable to higher holdings of liquid assets holdings among Malaysian banks (Graph 15) following the implementation of Basel III LCR requirement. On aggregate, Malaysian banks' liquid securities holdings stood at RM433 bil or 17.0% of total assets in 2017, a 32% increase since 2013.

To affirm that improved resilience at the aggregate system level is not down to few large banks alone, we also compared the breakdown of overall losses incurred by each bank over the same period (Graph 17). Consistent with the system level trends, fewer banks suffered from large losses, with only 2 banks suffering from losses exceeding 50% of CET1 capital. Again, during this period, bank capitalisation increased significantly (refer to Table 1) which could in part explain the improved resilience despite higher interconnectedness to cross-border financial institutions.

Banking System: Vulnerability by Type of Shocks and Round Graph 16



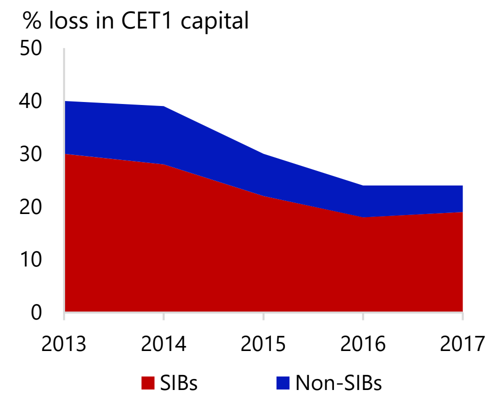
Breakdown of Losses (as % of CET1 capital) Graph 17



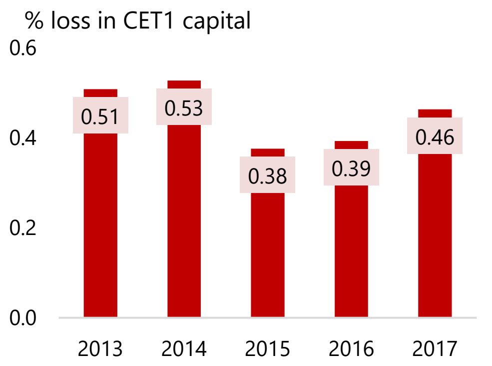
Source: Author's estimation

We then proceed to decompose aggregate losses across the five-year to losses attributable to SIBs and non-SIBs (Graph 18). Similarly, resilience of SIBs have improved over the years. Yet, unsurprisingly system-wide losses induced by SIBs remain significant. However, average vulnerability of SIBs, which reflects the average losses (as proportion of CET1 capital) incurred by SIBs across all iterations in Scenario 2, remain low (Graph 19).

Banking System: Vulnerability by Type of Shocks and Round — Graph 18



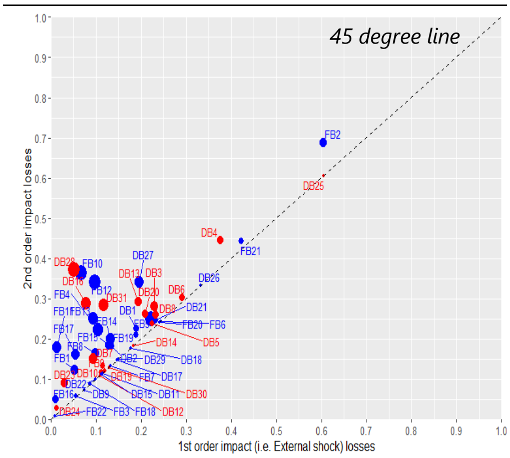
Breakdown of Losses (as % of CET1 capital) — Graph 19



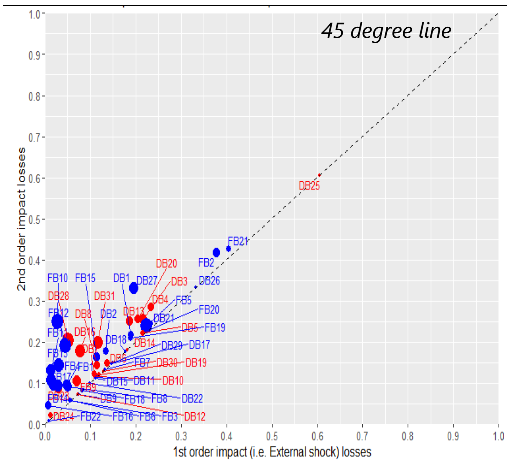
Source: Author's estimation

Next, we compare the first-round effects and the second-round effects (Graph 20). Here we observe banks with higher external and domestic interbank leverage suffer greater losses. These tend to be smaller foreign banks who typically operate as foreign currency liquidity provider in the domestic interbank market. As expected, overall impact is lower if related counterparties' exposures are excluded (Graph 21). Comparing with 2013 (Graph 22), we observe a significant reduction in the effects in 2017, with banks being clustered to the bottom left hand corner.

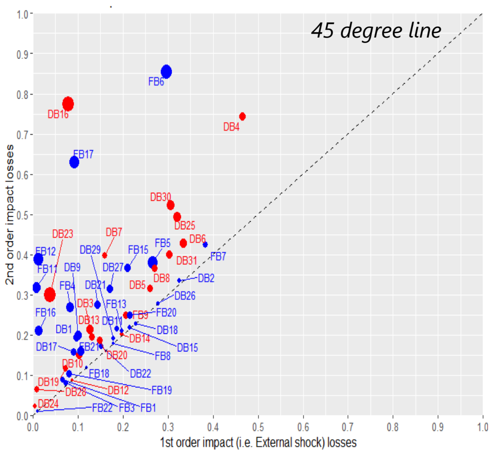
2017: First Order Impact vs Second Order Impact (as Proportion of CET1 Capital) — Graph 20



2017: First Order Impact vs Second Order Impact excluding Related Counterparties' Exposures (as Proportion of CET1 Capital) — Graph 21



2013: First Order Impact vs Second Order Impact (as Proportion of CET1 Capital) Graph 22



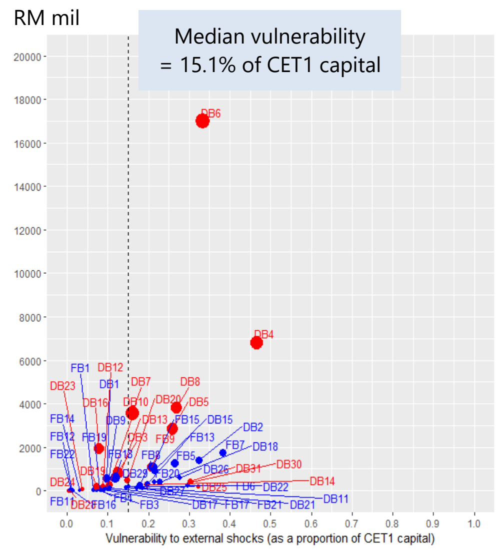
Note:

1. Marker size represents domestic interbank loans leverage (the bigger the marker, the greater the domestic interbank leverage)
2. Marker colour indicates type of bank: Red for SIBs, blue of non-SIBs

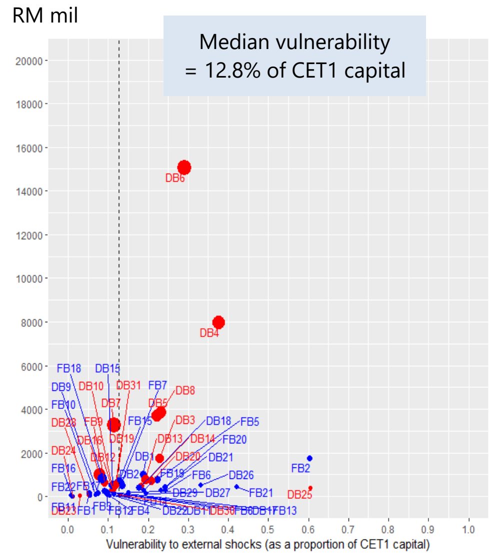
Source: Author’s estimation

Next, we focus on the two following quantities: (i) losses (as proportion of CET1 capital) incurred by a bank when faced with shocks in the first round and (ii) total loss induced by each bank onto the system when faced with shocks in the first round. The former, a measure of vulnerability to external shocks and the latter, a measure of impact a bank has onto the interbank market sheds light into two separate systemic risk aspects (Graphs 23 and 24). Comparing these quantities for 2013 and 2017, we observe that banks exhibiting relatively high external vulnerability are mostly small banks that also pose limited impact onto the domestic interbank market. Nevertheless, our results show that SIBs continue to exhibit high vulnerability (vulnerability exceeding median vulnerability) and can potentially pose significant impact due to their prominent role in the domestic interbank network. Unsurprisingly, size which is regarded as a key measure of systemic importance appear to be associated with high values of vulnerability and impact.

2013: Vulnerability to External Shocks vs Impact in Domestic Interbank Market Graph 23



2017: Vulnerability to External Shocks vs Impact in Domestic Interbank Market Graph 24



Note:

1. Marker size represents relative measure of bank assets (the bigger the marker, the larger the bank in terms of asset size)
2. Marker colour indicates types of bank: Red for SIBs, blue of non-SIBs

Source: Author's estimation

7. Conclusion and Policy Implications

We have demonstrated that interconnectivity and solvency contagion risk in the Malaysian banking system to external shocks have declined significantly since 2013, despite the rapid growth in cross-border exposures in value terms. Nonetheless, due to their large size and interconnectedness i.e. importance as a financial intermediator, SIBs are found to be more susceptible to induce contagion within the network. In assessing potential losses arising from contagion induced by external shocks, we incorporated two other channels of contagion that are often disregarded in similar counterfactual simulation studies: (i) valuation losses due to common asset exposures and (ii) mark-to-market losses resulting from credit quality deterioration or a loss of confidence (also known as 'credit quality channel', refer to Fink et.al (2016)).

We believe that network analysis and counterfactual simulation, particularly those which extends beyond simple spillover mechanisms, are good complements to other macroprudential surveillance tools that are being used to monitor financial system vulnerabilities. Such analysis helps to provide useful insights on financial system vulnerability to shock scenarios and the propensity of financial institutions to amplify contagion due to significance within a particular network and/or degree of interconnectedness to other financial institutions i.e. 'too connected to fail' which cannot be gauged by simply looking at the size of balance sheet. In this regard, our study has demonstrated that SIBs continue to exhibit relatively higher vulnerability to external shocks due to their cross-border exposures and can potentially cause significant impact onto Malaysian banking system during times of distress. This serves as a reminder on the need to be cognisant of the risks posed by systemically important banks and have necessary policy measures in place to both manage the risks and reduce the impact of SIBs failure. Furthermore, considering how financial interconnectedness is an increasingly regional and global phenomenon, there is a role for collective efforts by policymakers to improve the monitoring of systemic linkages. This should include having in place effective arrangements for multilateral information sharing and surveillance, and wider coverage of reporting on cross-border financial obligations.

Moving forward, this study could be extended in three key directions. First, with access to more complete information on banks cross-border exposures, particularly identity of counterparties, a more comprehensive and an accurate assessment of contagion risks posed by systemically important banks in other jurisdictions to the domestic financial system and economy can be conducted. Second, calibration of distress propagation mechanism for counterfactual simulation deserves further discussion. In this regard, a more refined and economically accurate mechanism that take into account counterparties' probability of default and the recovery rate of the assets could be explored (refer to Fink et. al., 2016; Bardoscia et.al, 2017). Third, our analysis can be expanded to assess contagion risk to take into account other form of interlinkages through which banks are interconnected with other financial institutions such as a wider set of asset classes e.g. cross holdings of securities or markets like OTC derivative markets. This would allow a more comprehensive assessment of contagion risk within the financial system arising from higher level of interconnectedness.

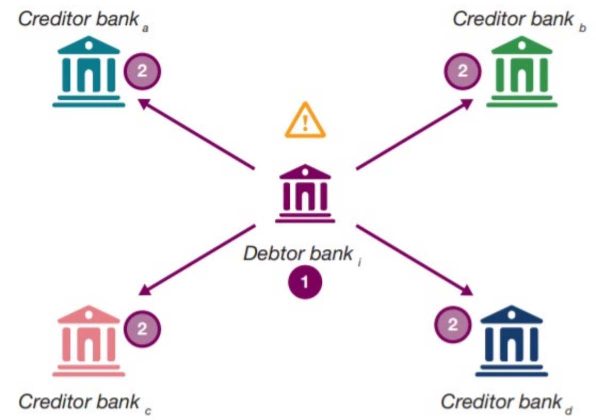
8. Addendum: Application of counterfactual simulation as part of the broader stress testing toolkit to assess contagion risk

This section is an extension to our earlier paper[23] and demonstrates the application of counterfactual simulation as part of broader stress testing toolkit to assess risk of contagion event in the domestic interbank market. Bank Negara Malaysia conducts a multi-year solvency stress test exercise on a regular basis to assess the resilience of banks under simulated scenarios of severe macroeconomic and financial strains.

The latest stress test exercise for the banking sector[24] incorporates the counterfactual simulation model studied in our paper to ascertain the incremental losses incurred by banks from potential contagion via domestic interbank market exposures. Here contagion is induced as banks incur losses from credit, market and external funding risk shocks under each adverse scenario of the macro stress test scenario (Diagram 1). A distressed bank's capacity to repay counterparties in the domestic interbank market is then assumed to decline as its creditworthiness deteriorates. Banks that lend to a distressed bank are then assumed to incorporate information about the reduced creditworthiness into the valuation of their interbank lending exposure to that distressed bank, thereby incurring losses. The losses incurred by the creditor bank is assumed to be proportionate to the relative loss of capital incurred by the debtor bank. For example, if the debtor bank incurs x% of loss in capital, the creditor bank is assumed to experience an equal relative loss of x% on their interbank lending exposure to the debtor bank.

Overall, the potential contagion losses arising from interbank exposures is assessed to be low at 6.3% of total losses in 2018. Unsurprisingly, losses from contagion risks are estimated to be lower if related counterparty exposures are excluded as most interbank market exposures are intragroup in nature.

Illustration of Contagion Risk in the Domestic Interbank Market — Diagram 1



Source: Bank Negara Malaysia

[23] See Harikumara and Lim (2019) "Cross-border Interbank Contagion Risk to the Malaysian Banking System".

[24] Refer to the Financial Stability and Payment Systems Report 2018, page 31-34 (Bank Negara Malaysia, 2019).

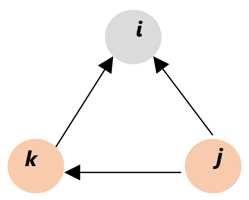
Appendices

Appendix 1: Network Data and Methodology

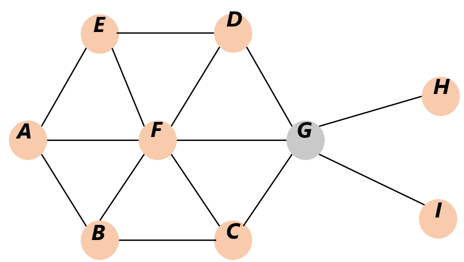
List of countries nodes — Table A1

No.	Country Node	Label	No.	Country Node	Label
1	Labuan	LABUAN	13	Japan	JPN
2	Singapore	SG	14	Republic of Korea	KOR
3	Cambodia	KHM	15	Taiwan	TWN
4	Indonesia	IDN	16	United States	US
5	Thailand	THA	17	United Kingdom	UK
6	Vietnam	VNM	18	Germany	DEU
7	Laos	LAOS	19	France	FRA
8	Brunei	BRN	20	Italy	ITA
9	Philippines	PHL	21	Spain	ESP
10	Myanmar	MMR	22	Other Asia	OTHER ASIA
11	Hong Kong SAR	HK	23	Other Europe	OTHER EU
12	China	CHN	24	Other Non-Residents	OTHER NR

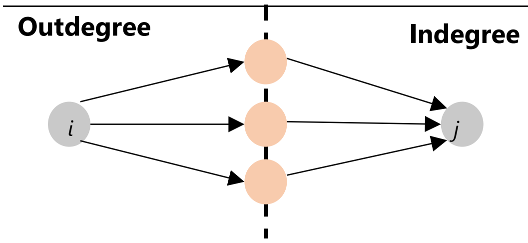
Clustering coefficient in a Core-Periphery Structure — Figure A1



Betweenness Centrality in a Network — Figure A2



Average Outdegree and Indegree — Figure A3



Appendix 2: Counterfactual Simulation Model

Description of the main variables		Table A2
Name	Symbol	Description
Individual bank vulnerability	hi(t)	Relative loss in equity of bank i at time t
System level vulnerability	H(t)	Relative loss of equity for the whole system at time t

Balance sheet dynamics — Table A3

Bank's assets are assumed marked-to-market whereas liabilities are carried at their face value

$$\text{Assets, } A_t \quad = \quad \text{Liabilities, } L + \text{Equity, } E_t \tag{4}$$

Each bank *i* either lend to other banks within the network i.e. domestic interbank lending (A_{dib}), or engage in any of the following (i) cross-border interbank lending (A_{ea}), (ii) placements with central bank (A_{cb}), (iii) holding of liquid securities (A_{liqs}) and (iv) holding of other assets (A_{oth}).

$$A_t^{dib} + A_t^{ea} + A_t^{cb} + A_t^{liqs} + A_t^{oth} \quad = \quad L + E_t \tag{5}$$

Total leverage of each bank at time t is the ratio of total assets and its equity, which can be disaggregated into its additive subcomponents as follows

$$\begin{aligned} \text{Total leverage, } l_t &= \frac{A_t}{E_t} \\ &= \frac{A_t^{dib}}{E_t} + \frac{A_t^{ea}}{E_t} + \frac{A_t^{cb}}{E_t} + \frac{A_t^{liqs}}{E_t} + \frac{A_t^{oth}}{E_t} \\ & l_t^{dib} + l_t^{ea} + l_t^{cb} + l_t^{liqs} + l_t^{oth} \end{aligned} \tag{9}$$

Contagion stress-test process	Table A4
At time, t =1 each bank experiences an effective negative credit shock, r on cross-border interbank lending exposures to bank k	
$\sum_k r_k A_{i,k}^{ea} = E_t \sum_k r_k l_{i,k}^{ea}$ *where* $A_i^{ea} = \sum_k A_{i,k}^{ea}$	(10)
The loss incurred on cross-border interbank lending results in corresponding reduction in equity for each bank i as follows:	
$A_{i,k,\, t=0}^{ea}$ - $A_{i,k,t=1}^{ea} = \sum r_k A_{i,k}^{ea} = E_{i,t=0} - E_{i,t=1}$	(11)
Simultaneously, each bank i also experiences a reversal shock, f on cross-border interbank borrowing exposures. Unable to replace the lost funding in the short-run, banks are forced to deleverage, first withdrawing their excess liquidity placements with central bank, Acb and then liquidating a portion, z of their holdings of liquid securities, Aliqs if there are insufficient placements with the central bank. A haircut, x on banks' remaining holdings of securities replicates the potential price impact of simultaneous asset liquidation to replace lost interbank funding and adverse market movements. The reduction in value of liquid securities results in corresponding reduction in equity is given by:	
$A_{i,\, t=0}^{liqs}$ - $A_{i,t=1}^{liqs} = x(1-z)A_{i,t=0}^{liqs} = E_{i,t=0} - E_{i,t=1}$	(12)
Cumulatively, individual relative equity loss at time t = 1 is computed as follows:	
Relative loss in equity, $h_{i,\, t=1} = \min\left\{1, \sum_k r_k A_{i,k}^{ea} + x(1-z) l_{i,t=0}^{liqs}\right\}$	(13)
Following this, the likelihood of a distressed bank (referring to bank suffering from relative loss in equity) repaying its obligations on the domestic interbank market becomes lower as its creditworthiness deteriorates. At time *t* =2, banks that lent to distressed banks will incorporate information about the reduced creditworthiness into the valuation of their interbank lending exposure thereby incurring losses. The value of the interbank asset will lie between its face value (repaid in full) and zero (assuming no recovery in the event of default). In our study, the losses incurred in the second round by creditor bank is proportional to relative loss of equity incurred by debtor banks in the first round. DebtRank excludes more than one reverberation, i.e. a bank can only transmit distress once to the network. We adopt similar premise in our study. Individual relative equity loss at time *t* = 2 is computed as follows:	
Relative loss in equity, $h_{i,\, t=2} = \min\left\{1, h_{i,t=1} + \sum_j l_{ij}^{dib} h_{j,t=1}\right\}$	(14)

Appendix 3: Results

Betweenness centrality (1Q 2013 vs 4Q 2017) — Table A2

SIB/Non-SIB	Betweenness Centrality		Change
	1Q 2013	4Q 2017	4Q 2017
SIB 4	0.03	0.08	↑
SIB 2	0.09	0.08	↓
SIB 1	0.08	0.08	↔
SIB 3	0.11	0.05	↓
NSIB 23	0.01	0.05	↑
SIB 4*	0.01	0.03	↑
NSIB 12	0.00	0.02	↑
NSIB 2	0.01	0.02	↑
SIB 5	0.04	0.02	↓
NSIB 1	0.09	0.01	↓

Note: *Subsidiaries of SIBs

Source: Author's estimations

References

Aoyama, H., Battiston, S. and Fujiwara, Y. (2013). DebtRank Analysis of the Japanese Credit Network. RIETI Discussion Paper Series 13-E-087. [online] Available at: https://www.rieti.go.jp/jp/publications/dp/13e087.pdf.

Bank Negara Malaysia (2013). Financial Stability and Payment Systems Report. [online] Bank Negara Malaysia, pp.46-51. Available at: http://www.bnm.gov.my/index.php?ch=en_publication&pg=en_fspr&ac=8&lang=en.

——— (2017). *Indebted to debt: An assessment of debt levels and financial buffers of households*. Financial Stability and Payment Systems Report. [online] Bank Negara Malaysia, pp.39-48. Available at: http://www.bnm.gov.my/files/publication/fsps/en/2017/cp01_001_box.pdf

——— (2018). Financial Stability Review - First Half 2018. [online] Bank Negara Malaysia, pp.20-24. Available at: http://www.bnm.gov.my/files/publication/fsreview/en/2018/en_fs_01.pdf

——— (2019). Financial Stability and Payment Systems Report 2018. [online] Bank Negara Malaysia, pp.31-34. Available at: http://www.bnm.gov.my/files/publication/fsps/en/2018/fs2018_book.pdf

Bardoscia, M., Battiston, S., Caccioli, F. and Caldarelli, G. (2015). DebtRank: A Microscopic Foundation for Shock Propagation. *PLOS ONE*, 10(6), p.e0130406.

Bardoscia, M., Caccioli, F., Perotti, J., Vivaldo, G. and Caldarelli, G. (2016). Distress Propagation in Complex Networks: The Case of Non-Linear DebtRank. PLOS ONE, 11(10), p.e0163825.

Bardoscia, M., Barucca, P., Codd, A. and Hill, J. (2017). *The decline of solvency contagion risk*. Staff Working Paper No. 662. [online] Available at: https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2017/the-decline-of-solvency-contagion-risk [Accessed 7 Jan. 2019].

Basel Committee on Banking Supervision (2014). *Impact and implementation challenges of the Basel framework for emerging market, developing and small economies*. Working Paper No 27. [online] Available at: https://www.bis.org/bcbs/publ/wp27.pdf [Accessed 7 Jan. 2019].

———. (2018). *Capital treatment for bilateral counterparty credit risk finalised by the Basel Committee*. [online] Available at: https://www.bis.org/press/p110601.htm [Accessed 17 Dec. 2018].

Battiston, S., Delli Gatti, D., Gallegati, M., Greenwald, B. and Stiglitz, J. (2012). Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *Journal of Economic Dynamics and Control*, 36(8), pp.1121-1141.

Battiston, S., Puliga, M., Kaushik, R., Tasca, P. and Caldarelli, G. (2012). DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk. *Scientific Reports*, 2(1).

Battiston, S., Caldarelli, G., D'Errico, M. and Gurciullo, S. (2015). Leveraging the Network: A Stress-Test Framework Based on DebtRank. SSRN Electronic Journal.

Burrows, O., Learmonth, D., McKeown, J. and Williams, R. (2012). RAMSI : a top-down stress-testing model developed at the Bank of England. Quarterly Bulletin 2012 Q3. [online] Bank of England, pp.204-210. Available at: 7.

https://www.bankofengland.co.uk/-/media/boe/files/quarterly-bulletin/2012/ramsi-a-top-down-stress-testing-model-developed-at-the-boe.pdf [Accessed 28 Dec. 2018].

Elsinger, H., Lehar, A. and Summer, M. (2012). Network models and systemic risk assessment. [online] Available at: https://www.researchgate.net/publication/309187281_Network_models_and_systemic_risk_assessment [Accessed 7 Jan. 2019].

Fink, K., Krüger, U., Meller, B. and Wong, L. (2016). The credit quality channel: Modeling contagion in the interbank market. *Journal of Financial Stability*, 25, pp.83-97.

Glasserman, P. and Young, H. (2015). How likely is contagion in financial networks?. *Journal of Banking & Finance*, 50, pp.383-399.

Hattori, M. and Suda, Y. (2007). Developments in a cross-border bank exposure network. *Bank of Japan Working Paper*, No. 07–E–21.

Harikumara. S and Lim. S.L. (2019). Cross-border Interbank Contagion Risk in the Malaysian Banking System. *SEACEN Cross-border Interbank Contagion Risk Analysis*, pp.109-141.

International Monetary Fund (IMF), Bank for International Settlements (BIS), the Commonwealth Secretariat (Comsec), Eurostat, the Organization for Economic Co-operation and Development (OECD), the Paris Club Secretariat, the United Nations Conference on Trade and Development (UNCTAD) and the World Bank. (2013). External Debt Statistics: Guide for Compilers and Users. Available at: https://unstats.un.org/unsd/nationalaccount/docs/External_debt_statistics_guide.pdf [Accessed 28 Dec. 2018].

Kubelec, C. and Sa, F. (2010). The Geographical Composition of National External Balance Sheets: 1980-2005. *SSRN Electronic Journal*.

OECD (2012)., "Financial Contagion in the Era of Globalised Banking?.", OECD Economics Department Policy Notes, No. 14, June

Opsahl, T., Agneessens, F. and Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks, 32(3), pp.245-251.

Remolona, E. and Shim, I. (2015). The rise of regional banking in Asia and the Pacific. BIS Quarterly Review, September 2015. [online] pp.119-134. Available at: https://www.bis.org/publ/qtrpdf/r_qt1509j.pdf [Accessed 28 Dec. 2018].

Reyes, J. and Minoiu, C. (2011). A Network Analysis of Global Banking:1978-2009. *IMF Working Papers*, 11(74), p.1.

Summer, M. (2013). Financial Contagion and Network Analysis. *Annual Review of Financial Economics*, 5(1), pp.277-297.

SEACEN (2019). Cross-border Interbank Contagion Risk Analysis. *SEACEN Research Publications*.

Tabak, B., Souza, S. and Guerra, S. (2013). Assessing Systemic Risk in the Brazilian Interbank Market. Banco Central do Brasil Working Papers, [online] Working Paper n. 318. Available at: https://www.bcb.gov.br/pec/wps/ingl/wps318.pdf [Accessed 17 Dec. 2018].

Upper, C. (2011). Simulation methods to assess the danger of contagion in interbank markets. Journal of Financial Stability, 7(3), pp.111-125.

62nd ISI WORLD STATISTICS CONGRESS 2019

18 – 23 August 2019, Kuala Lumpur

Come | Connect | Create

IFC-BNM-ECB Satellite Seminar
on "New analytics for working with micro data, AI"

Cross-border contagion risk to the Malaysian banking system

Harikumara Sababathy & Lim Sheng Ling
Bank Negara Malaysia

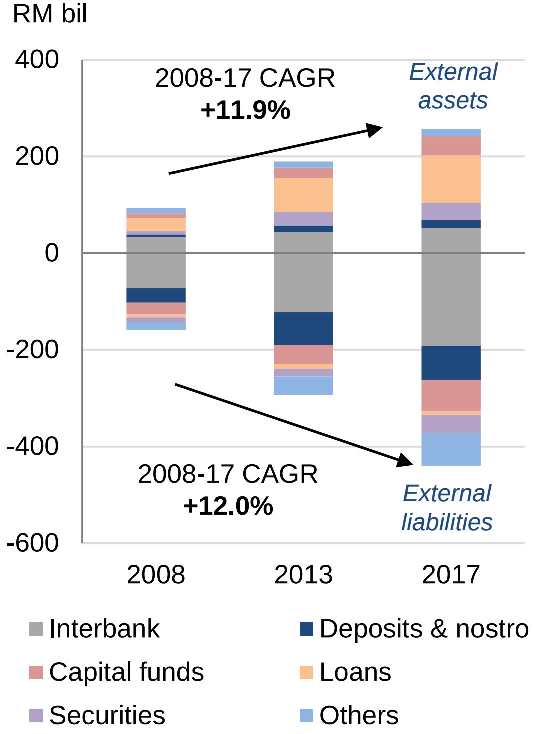
16th August 2019

Presentation outline

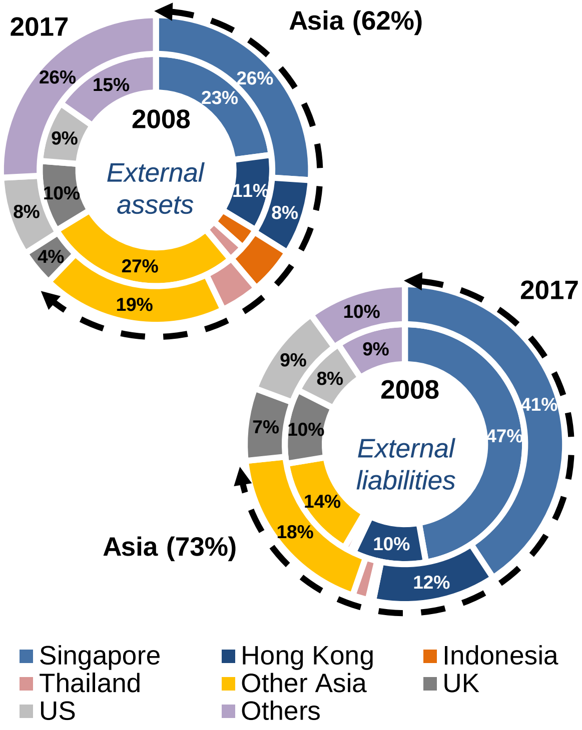
1. Setting the scene
2. Research objectives
3. Data
4. Methodologies
   - Network analysis & visualisation
   - Contagion stress test
5. Results & policy implications

Why are we concerned about cross-border contagion risk?

Malaysian banks' cross-border exposures has expanded rapidly in recent years…



…with significant concentration in the Asian region



This is a natural consequence of the Malaysian banking structure

1 Malaysian banking system is both concentrated and large

Total assets **231%** of **GDP**

Top-5 banks **70%** of total assets

2 Sizeable presence of foreign banks in banking sector

20% of total assets

3 Strong overseas presence of domestic banks regionally in Asia

Total foreign claims **49%** of **GDP**

Source: Bank Negara Malaysia, BIS

Note: All data based on 2017

Research objectives

1. Explore properties of Malaysian banks' interbank exposures using network analysis techniques, including assessment of dynamics over time
2. Develop a cross-border contagion stress-test that will facilitate the following:
   - Assessment of banks' susceptibility to external shocks at the macro- and micro- level
   - Estimation of contagion path (propagation of shocks) across banks via the domestic interbank market induced by external shocks and potentially vice-versa
   - Identification of potential systemic linkages and cross-border domino effects between systemically-important banks (SIBs), internationally active banks and the Malaysian banking system

Data

<table>
<tr><th>Methodology</th><th colspan="2">(i) Network analysis</th><th>(ii) Counterfactual simulation</th></tr>
<tr><td rowspan="2">Source</td><td colspan="3">• Banks’ bilateral exposures in the domestic interbank market (Bank Negara Malaysia, BNM)
• Banks’ cross-border exposures (BNM, aligned to BIS international banking statistics)</td></tr>
<tr><td colspan="2"></td><td>• Regulatory capital (BNM)
• Placements with central bank (BNM)
• Holdings of liquid securities (BNM)</td></tr>
<tr><td>Frequency</td><td colspan="2">Quarterly (2013-2017)</td><td>Annual (2013-2017)</td></tr>
<tr><td rowspan="2">Number of observations</td><td rowspan="2">78 nodes</td><td>Core: 54 banks</td><td rowspan="2">54 banks</td></tr>
<tr><td>Periphery: 24 external counterparty countries</td></tr>
<tr><td></td><td colspan="2">• Does not consider the possibility of new links
• Only positive flows (net placements/net assets) are retained
  o Negative flows (net liabilities) are zerorised</td><td>(further details in subsequent slides)</td></tr>
</table>

Note:

1. In this study, we have identified 20 out of 54 banks, comprising domestic and foreign banks, as systemically important banks to facilitate our analyses. On a consolidated basis, these SIBs are part of 7 banking groups. Such banks shall be referred to as systemically important banks (SIBs) hereafter.
2. SIBs refers to banks whose distress or failure may result in severe negative spillover impact to the financial system and real economy. At the global stage, the Financial Stability Board and Basel Committee of Banking Supervision assess and identify banks that are deemed to be systemically important on an annual basis, referred to as global systemically important banks (G-SIB). Similarly, national authorities identify banks that are systemically important in their respective jurisdiction, commonly referred to as domestic systemically important banks (D-SIB).
3. SIBs for our study are selected based on a combination of balance sheet indicators namely total assets, total deposits and total loans outstanding.

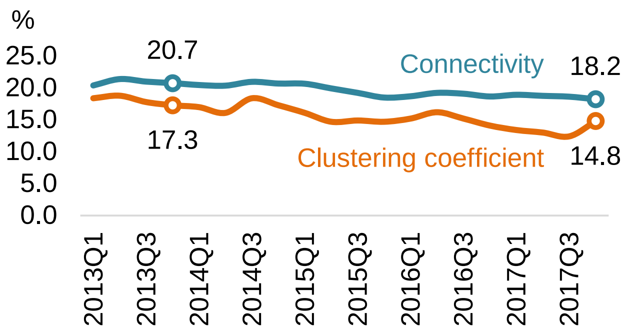
Assessing Topological Characteristics of Networks

1. **Connectivity** — Likelihood of connection between nodes
2. **Clustering Coefficient** — Likelihood of connection between neighbours
3. **Betweenness Centrality** — The number of times a node lies on the shortest path between other nodes
4. **Average Degree** — Ratio of total number of links to total number of nodes

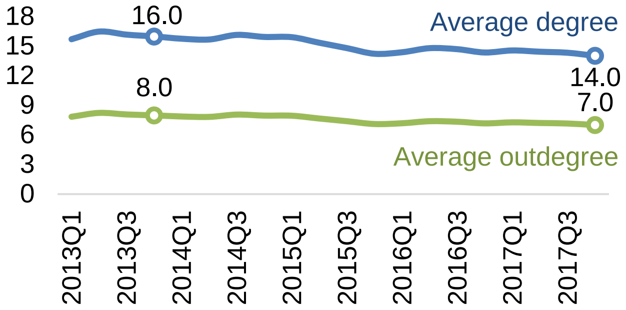
Connectivity of banks in the domestic interbank market decreases over time

18.2% of banks are likely to be connected in 2017 (2013: 20.7%)

Connectivity and clustering coefficient



Average Degree and Outdegree



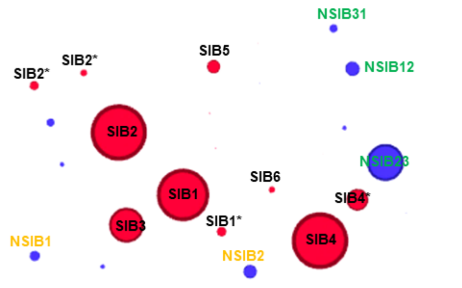
Source: Bank Negara Malaysia

Nonetheless, the role of SIBs as key financial intermediaries remain important

SIBs: Betweenness centrality (1Q 2013 vs 4Q 2017)

SIB/ Non-SIB	Betweenness Centrality		Change
	1Q 2013	4Q 2017	
SIB 4	0.03	0.08	↑
SIB 2	0.09	0.08	↓
SIB 1	0.08	0.08	↔
SIB 3	0.11	0.05	↓
Non-SIB 23	0.01	0.05	↑
SIB 4*	0.01	0.03	↑
Non-SIB 12	0.00	0.02	↑
Non-SIB 2	0.01	0.02	↑
SIB 5	0.04	0.02	↓
Non-SIB 1	0.09	0.01	↓

4Q 2017 Network: Betweenness centrality

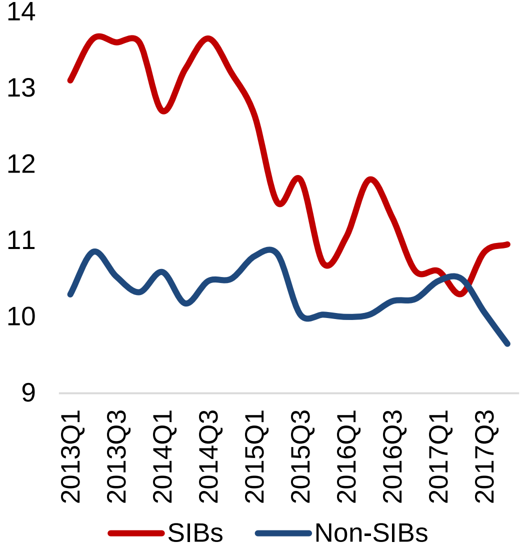


Note: * indicates subsidiaries or affiliates of the same banking group

Increase in intensity of cross-border flows by SIBs reduces outdegree

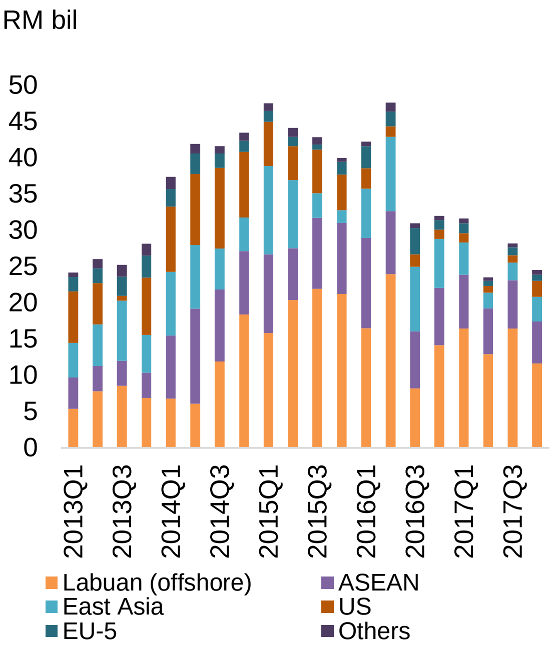
SIBs cross-border exposures has increased…

Banking System: Average Outdegree



…with greater exposure to banks in Asian region…

SIBs: Net interbank exposures to non-residents by region



… amid higher intragroup exposures to non-residents

SIBS: Net intragroup exposure to non-residents

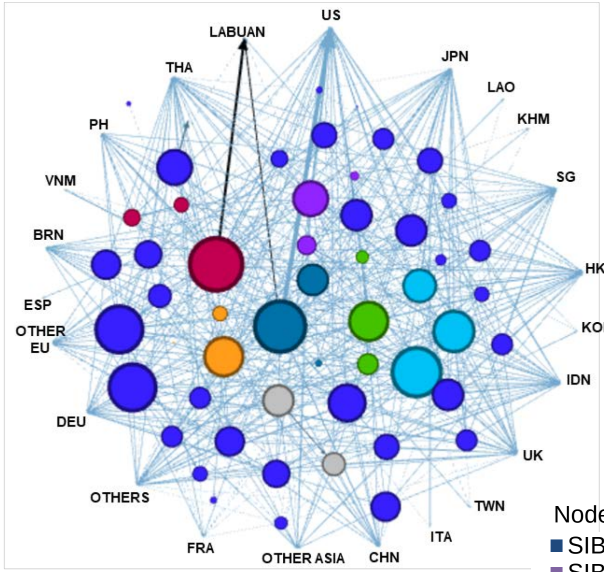
Source: Bank Negara Malaysia

Decreasing connectivity in the Malaysian banks' interbank market

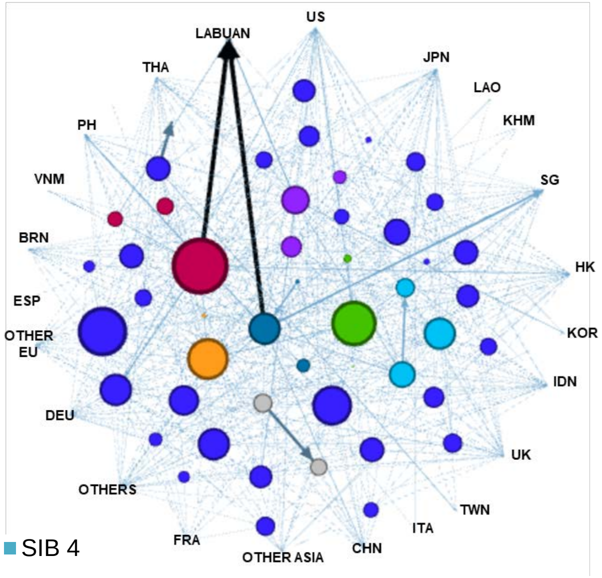
Key observations:

- **Interbank activities are concentrated among intragroup entities**
- **Increase of high quality liquid assets (central bank placements and securities holdings) to meet Liquidity Coverage Ratio (LCR) requirements**

1Q 2013: Network outdegree



4Q 2017: Network outdegree



Nodes legend:
■ SIB 1 ■ SIB 2 ■ SIB 3 ■ SIB 4
■ SIB 5 ■ SIB 6 ■ SIB 7

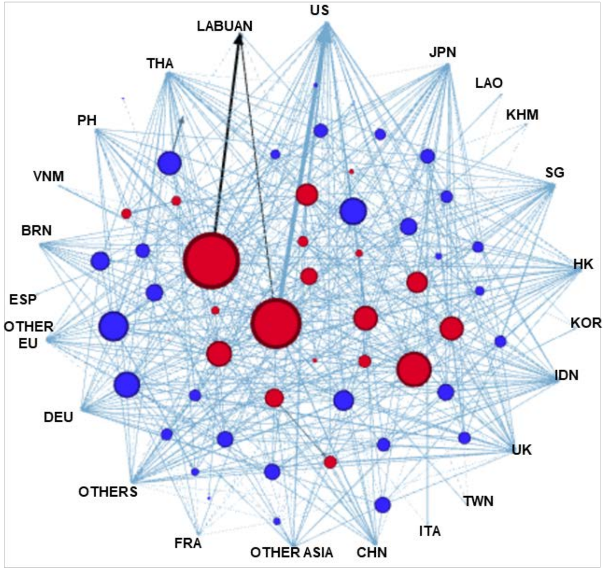
Note: - Royal blue nodes represent non-SIBs. Other coloured nodes represent SIBs on a consolidated basis (7 in total).
- Size of the nodes represent the number of counterparties
- Direction and thickness of the arrows represent net placement and relative size of exposures

Source: Bank Negara Malaysia

However, more outgoing links does not always equal greater contagion risk

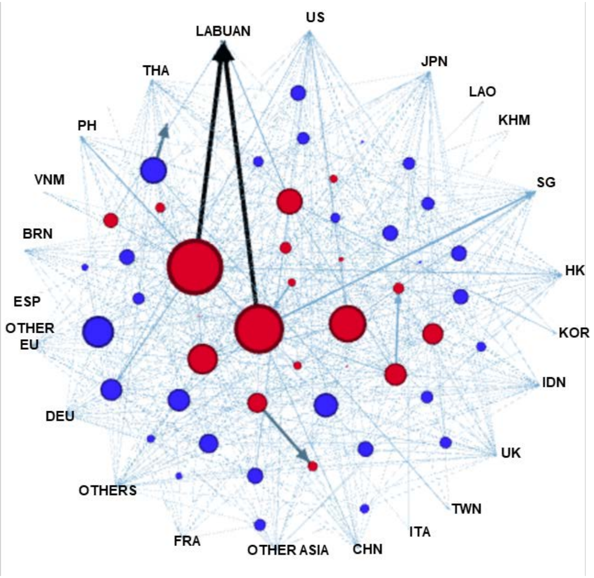
DSIBs are more susceptible to contagion risk compared to non-DSIBs…

1Q 2013: Weighted network outdegree



…but the risk for some banks decreased over time

4Q 2017: Weighted network outdegree



Note: - Red nodes represent SIBs, blue nodes represent non-SIBs
- Size of the nodes represent the degree of interconnectedness, taking into account number of counterparties and size of exposures
- Direction and thickness of the arrows represent net placement and relative size of exposures

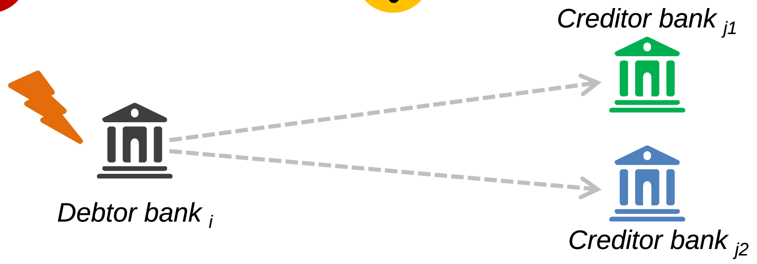
Source: Bank Negara Malaysia

Assessing contagion risk arising from external exposures

Methodology

Shocks

Domestic interbank market



First order impact

- Banks experience losses due to credit and funding shocks on their external exposures, and are deemed to be distressed even if not in default.
- The likelihood of a distressed bank repaying its obligations becomes lower as its creditworthiness deteriorates.

Second order impact

- The banks that lent to a distressed bank will incorporate information about the reduced creditworthiness into the valuation of their interbank lending exposure to the that bank thereby incurring losses.

[1] Based on 'DebtRank' methodology developed by Battison et. al (2012) in "DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk" to capture distress propagation in networks without assuming defaults

Scenario & parameters

Credit shock

50% default on cross-border interbank lending

Funding & market shocks

50% withdrawal of short-term external debt with unrelated counterparties, lost funding is not refinanced

5% haircut on securities holdings to simulate impact of simultaneous asset liquidation to replace lost funding & adverse market movements

Scenarios

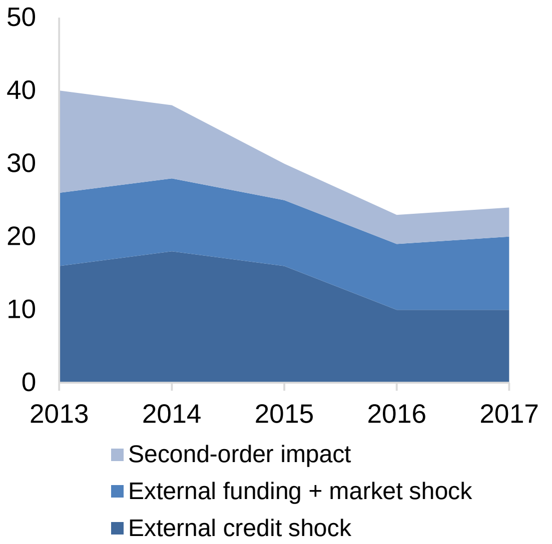
1. Common shock affecting all banks simultaneously at time *t=1*
2. A single bank is shocked at time *t=1*. This is repeated for each bank in the system

Banks' resilience to external shocks have improved over time

Consistent with lower connectivity in the interbank market, second order impact while sizeable has declined

Banking System: Vulnerability by Type of Shocks and Round

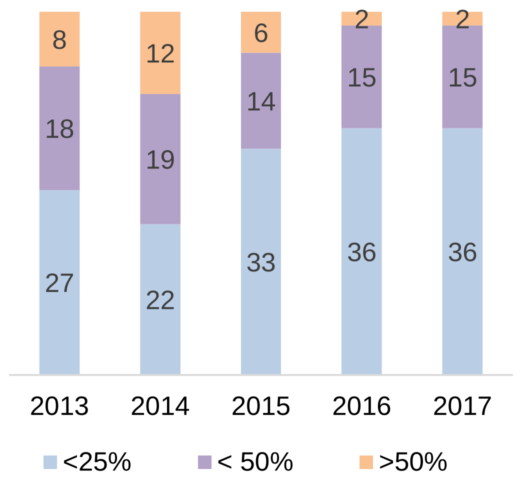
% loss in CET1 capital



Fewer banks suffered from large losses, with only 2 banks significant losses

Breakdown of Losses (as % of CET1 capital)

Number of banks



Similarly, resilience of SIBs has improved during the same period

1 While relative capital losses induced by SIBs remains significant…

79% of total losses

2 …average vulnerability[1] of SIBs are low

0.46% of CET1 capital

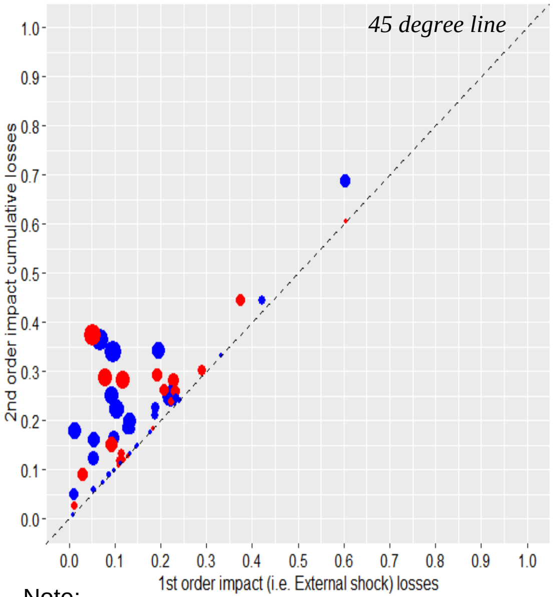
Note:

[1]Average vulnerability reflects the average losses (as proportion of CET1 capital) incurred by a bank across all iterations when each bank in the system is shocked individually (Scenario 2)

Banks with higher domestic interbank leverage suffer greater losses

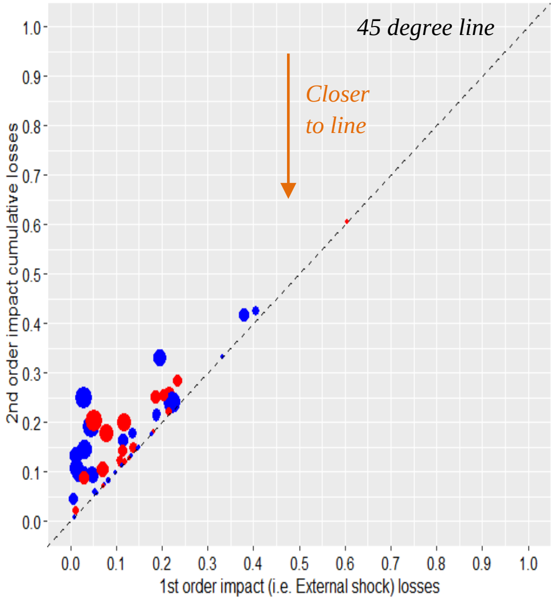
These are largely foreign banks that operate as FCY liquidity provider (and have smaller capital base)

2017: 1st Order Impact vs 2nd Order Impact (% of CET1 Capital)



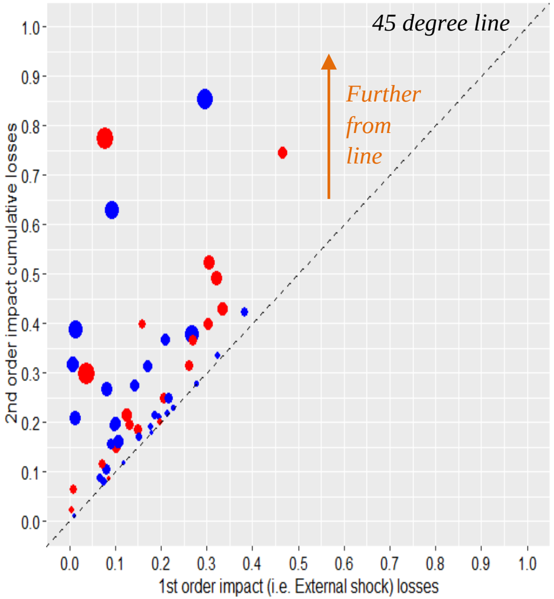
However, overall impact is lower if related counterparties' exposures are excluded

2017: Excluding Related Counterparties' Exposures



Since, 2013, there has been reduction in the cumulative impact of external shocks

2013: 1st Order Impact vs 2nd Order Impact (% of CET1 Capital)

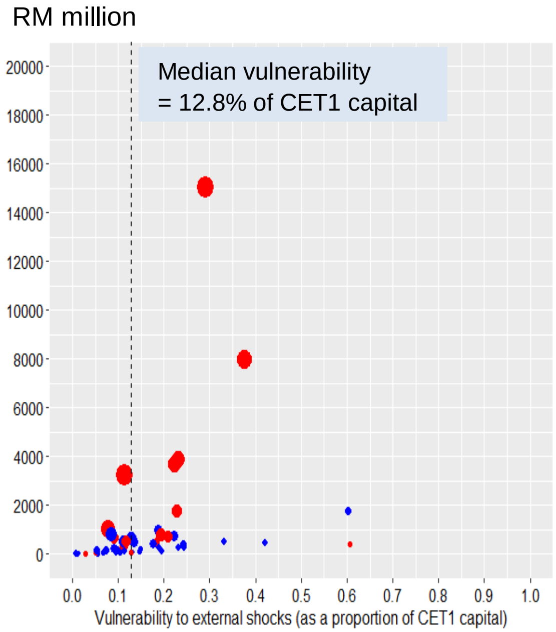


Note:

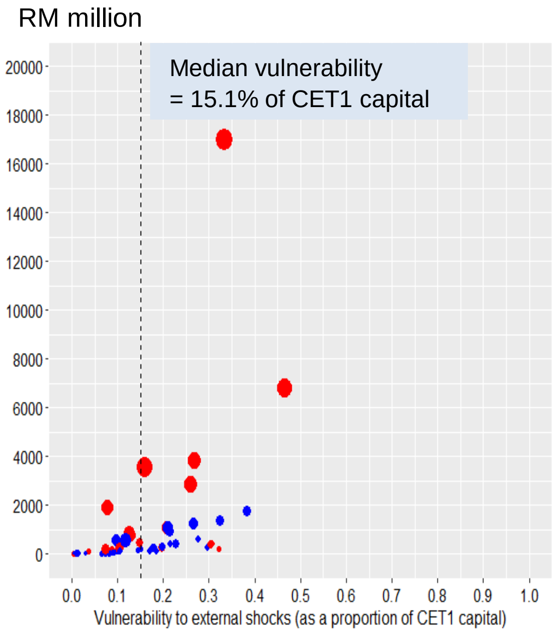
1. Marker size represents domestic interbank loans leverage (the bigger the marker, the greater the domestic interbank leverage)
2. Marker colour indicates type of bank: Red for SIBs, blue of non-SIBs

SIBs cause significant impact onto system

2017: Vulnerability to External Shocks vs Impact in Domestic Interbank Market



2013: Vulnerability to External Shocks vs Impact in Domestic Interbank Market



- Banks exhibiting high external vulnerability are mostly small banks that also pose limited impact onto the domestic interbank market
- However, SIBs continue to exhibit relatively higher vulnerability and cause significant impact onto system
- Unsurprisingly, size (larger dots) appears to be associated with high values of individual vulnerability and systemic impact

Note:

1. Marker size represents represents relative measure of bank assets (the bigger the marker, the larger the bank in terms of asset size)
2. Marker colour indicates type of bank: Red for SIBs, blue of non-SIBs

Conclusion & policy implications

Results

1. **Interconnectivity in the domestic interbank market and solvency contagion risk of Malaysian banks to external shocks have declined significantly since 2013**

2. **Nonetheless, due to their large size and interconnectedness i.e. importance as a financial intermediator, SIBs are found to be more susceptible to induce contagion within the network, causing significant impact onto the system**

Policy implications

1. **Network analysis and counterfactual simulation are good complements to other macroprudential surveillance tools that are being used to monitor financial system vulnerabilities**
   - Such analysis helps to provide useful insights i.e. 'too connected to fail' which cannot be gauged by simply looking at the size of balance sheet

2. **Importance of comprehensive and effective arrangements in place for crisis containment, management and resolution**
   - Impact of distress or failure of SIBs will impose outsized impact on the system

3. **Importance of having effective arrangements for multilateral information sharing and surveillance**

62nd ISI World Statistics Congress
ISI 2019 Kuala Lumpur
18 - 23 August 2019 | www.isi2019.org

THANK YOU

Harikumara Sababathy (harikumara@bnm.gov.my) &
Lim Sheng Ling (sheng@bnm.gov.my)

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Predicting and preventing financial crisis- where do we stand?[1]

Boris Hofmann,

Bank for International Settlements

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Predicting and preventing financial crises: Where do we stand?

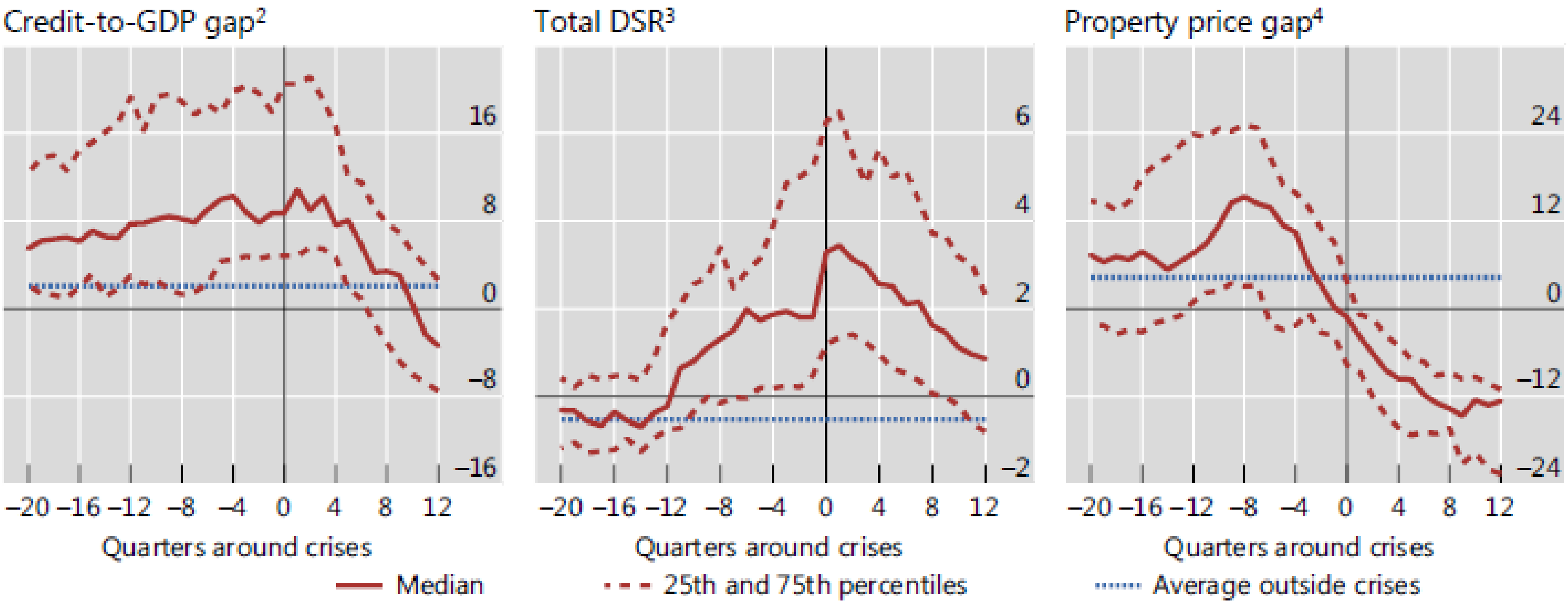
Boris Hofmann (BIS)

IFC-BNM-ECB Satellite Seminar "Post-crisis data landscape: micro data for the macro world“, Bank Negara Malaysia, Kuala Lumpur, 16 August 2019

Outline

- Predicting financial crises
- Preventing financial crises
- Special challenges for EMEs

Aggregate crisis indicators…



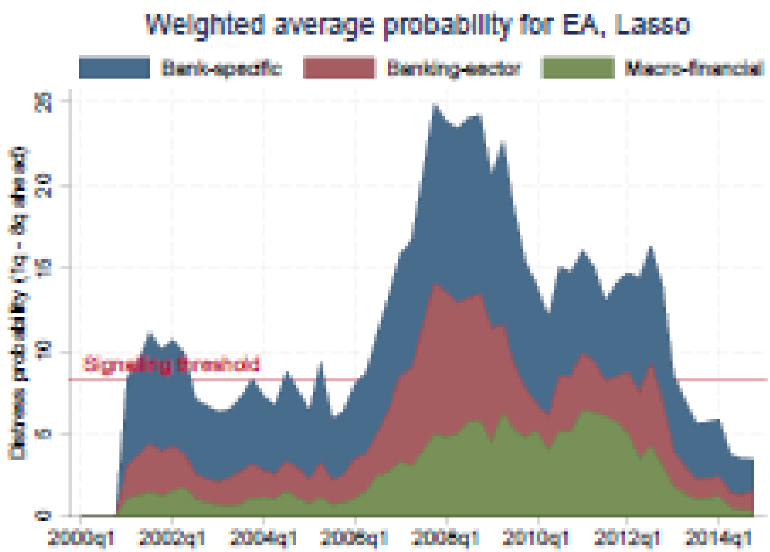
…with predictive power at different horizons (AUC)

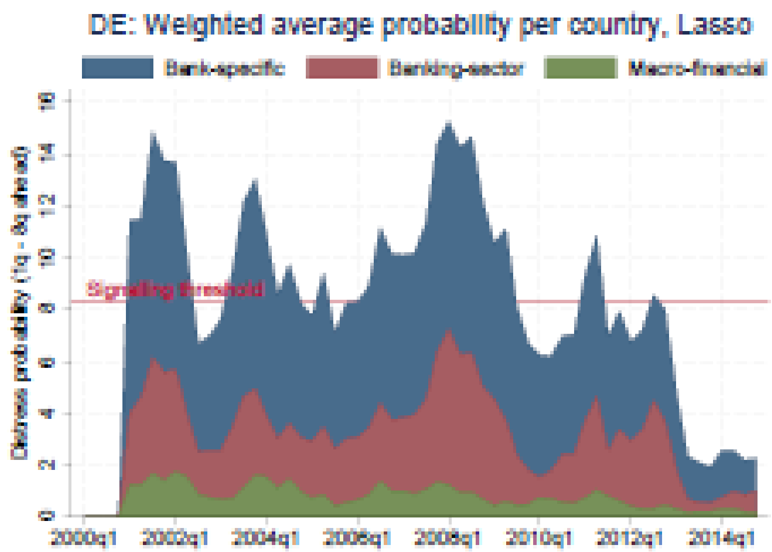
EWI	Horizon (quarters)											
	1	2	3	4	5	6	7	8	9	10	11	12
Credit-to-GDP gap	0.80*	0.80*	0.81*	**0.80***	**0.79***	**0.78***	**0.75***	**0.78***	**0.77***	**0.78***	**0.76***	**0.77***
Total DSR	**0.84***	**0.83***	**0.82***	0.79*	0.77*	0.76*	0.74*	0.75*	0.74*	0.71*	0.69*	0.69*
Property price	0.43	0.49	0.54	0.61	0.64	0.67*	0.68*	0.72*	0.71*	0.70*	0.69*	0.68*
Household DSR	0.82*	0.81*	0.80*	0.79*	0.77*	0.76*	0.73*	0.76*	0.75*	0.72*	0.69*	0.67*
Household credit-to-GDP gap	0.60	0.60	0.61	0.65*	0.66*	0.67*	0.70*	0.74*	0.75*	0.76*	0.76*	0.76*
Foreign currency debt to GDP	0.73*	0.71*	0.71*	0.67*	0.64	0.60	0.52	0.47	0.49	0.46	0.46	0.45
Cross-border claims to GDP	0.75*	0.75*	0.77*	0.74*	0.73*	0.71*	0.67*	0.66*	0.65*	0.63*	0.60	0.58

Source: Aldasoro et al (2018)

Micro data and financial crisis prediction

- Micro data can be used to predict risk of individual bank distress (Lang et al (2018))
  - Relate observed cases of bank distress to bank-level data as well as aggregate banking sector and macro-financial variables
- Aggregation of bank distress risk yields measure of systemic risk
  - Weighted average of bank distress probabilities





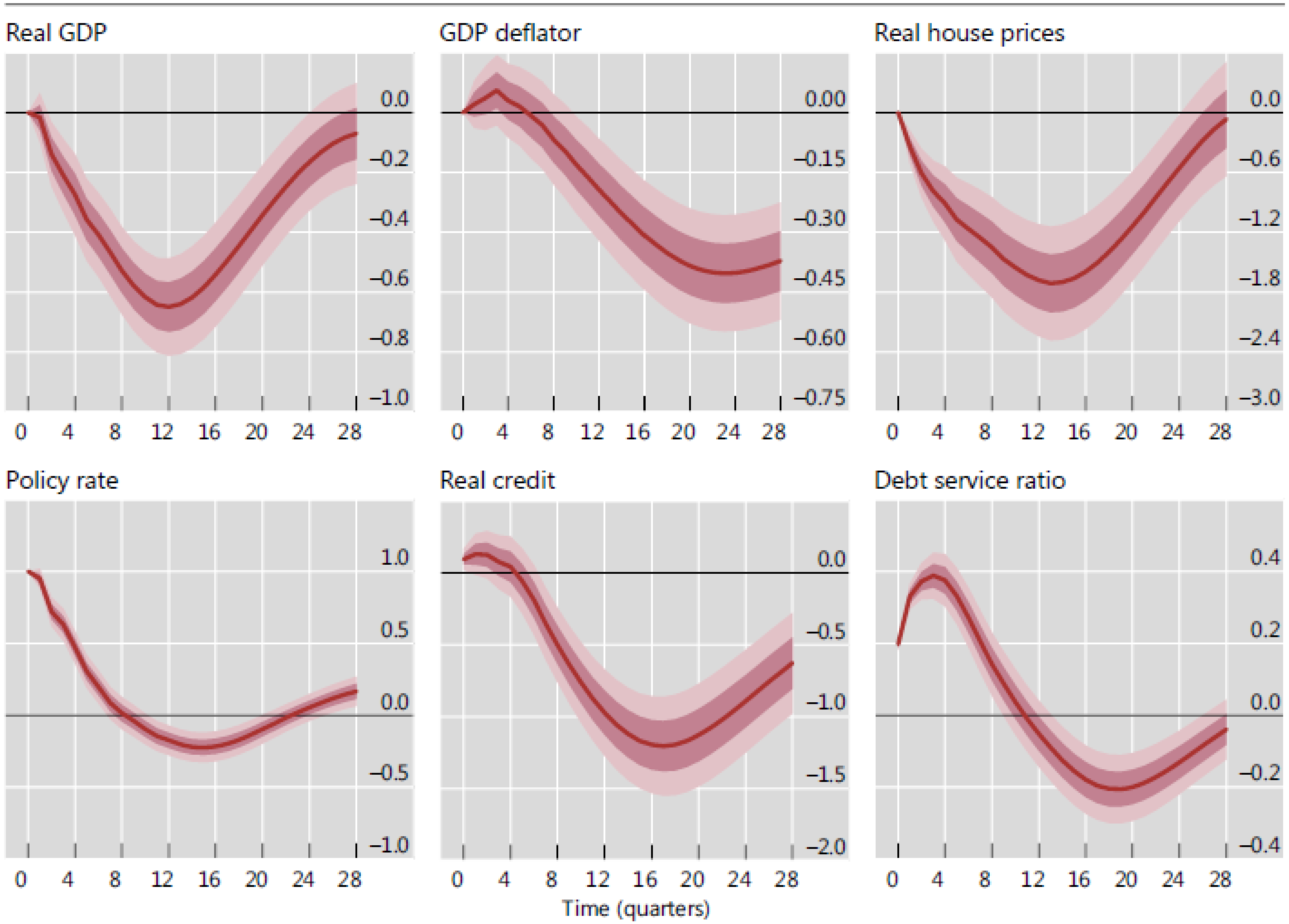
Source: Lang et al (2018)

Preventing financial crises

- Macroprudential policy
  - Countercyclical RR, LTV/DTI caps
  - Address financial imbalances in a targeted manner
    - Less collateral damage, but more prone to circumvention
    - Macropru enhances crisis resilience (BIS (2018))

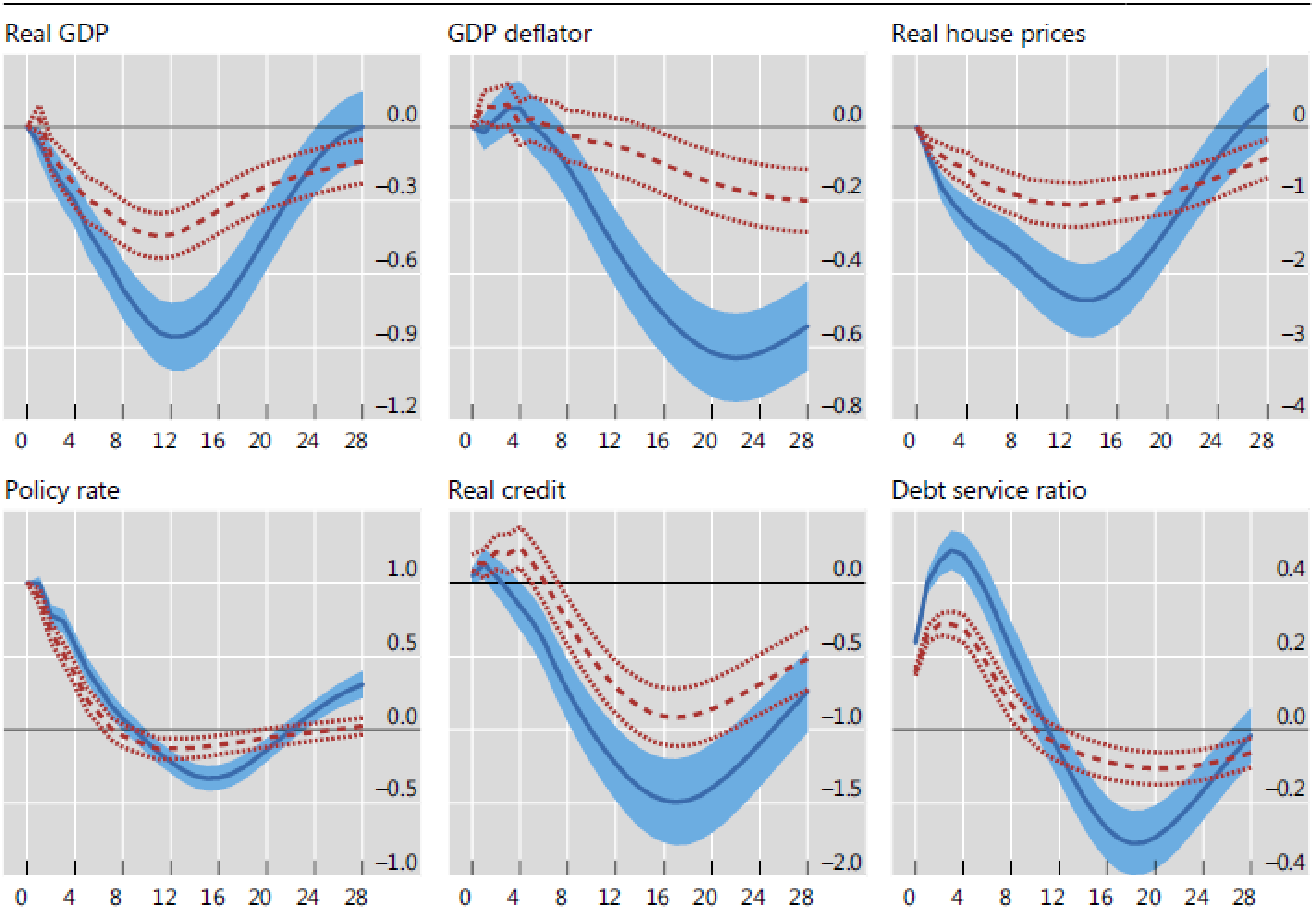
- Monetary policy
  - Tighter monetary policy to lean against financial imbalances
  - "Gets into all of the cracks"
  - But "collateral damage" on real economy
    - Credit over GDP might rise
    - Debt service burden might rise

Macro-financial impact of monetary tightening: Panel evidence



Source: Hofmann and Peersman (2017)

The role of debt levels



Source: Hofmann and Peersman (2017)

Evidence from micro data

- Microdata facilitate the identification of loan demand and supply factors and hence of the effects of macropru and monetary policy

- Analysis of the effects of macropru and monetary policy using bank-loan data from credit registries in Latin American countries (meta-analysis in Gambacorta and Murcia Pabon (2017))
  - Macropru and monetary policy are complementary: greater effectiveness when both policies push in the same direction

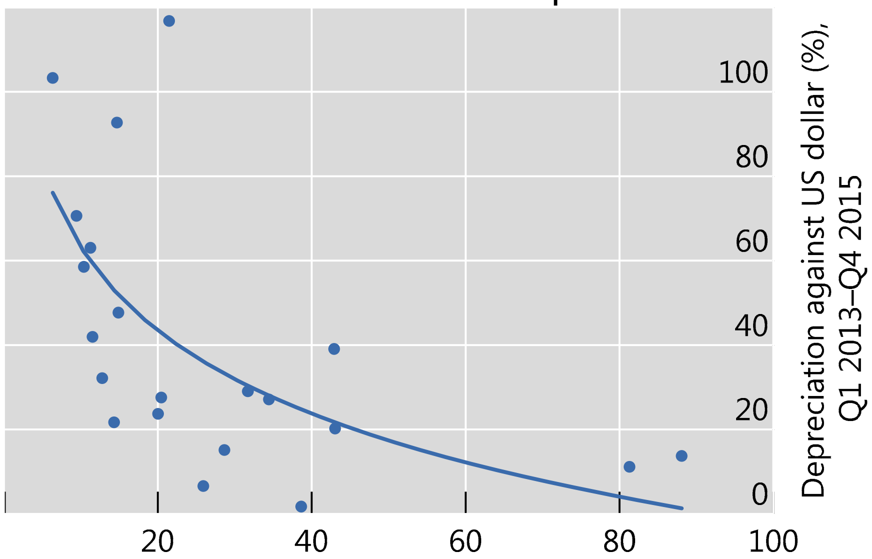
Special challenges for EMEs

- Dependence on foreign funding and less developed financial systems raise vulnerability to global financial factors

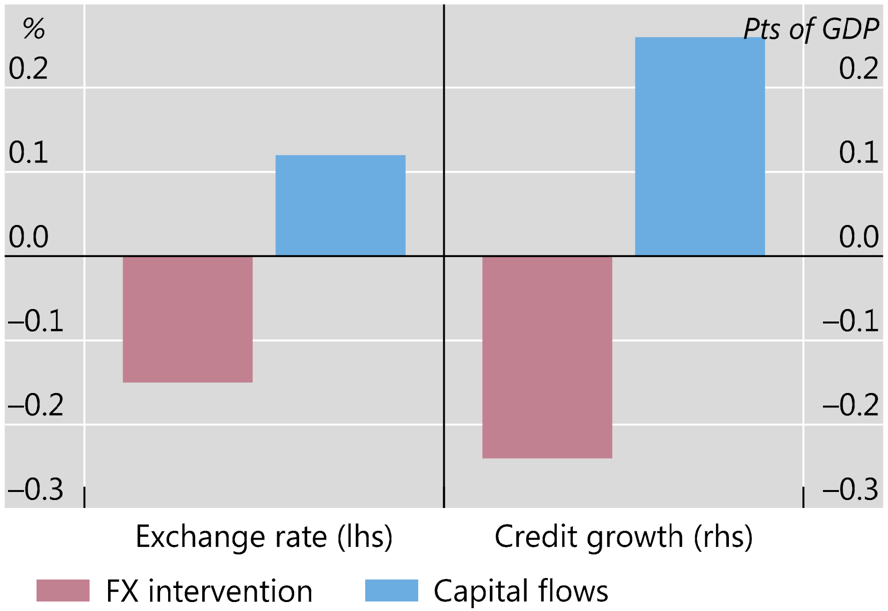
- Borrowing in foreign currency or in local currency from foreign lenders gives rise to a financial channel of the exchange rate
  - Capital inflows appreciate the exchange rate, which reduces credit risk in the presence of currency mismatches
  - Mutually reinforcing feedback loop between capital inflows and exchange rate appreciation
  - BIS AER 2019 special chapter on monetary policy frameworks in EMEs

FX intervention enhances resilience: Aggregate evidence

FX reserves cushion shock impacts

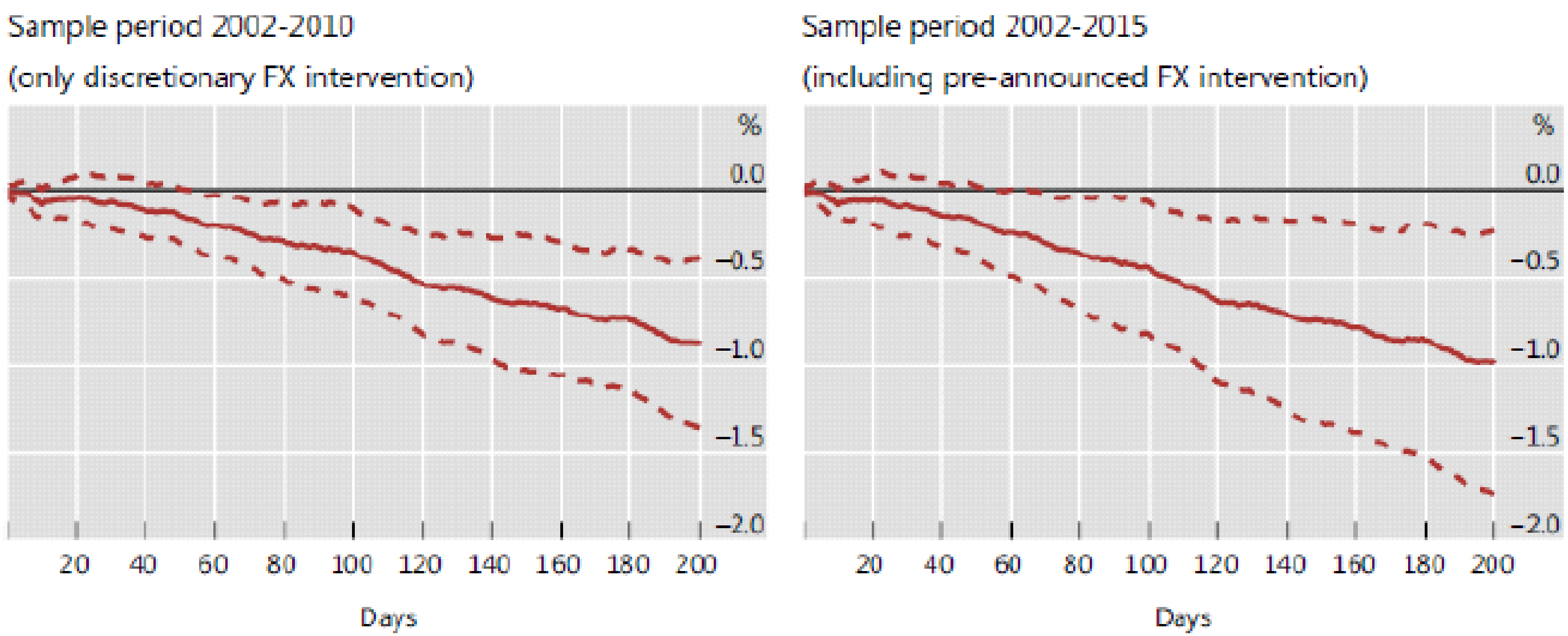


FX intervention effects



Source: BIS (2019)

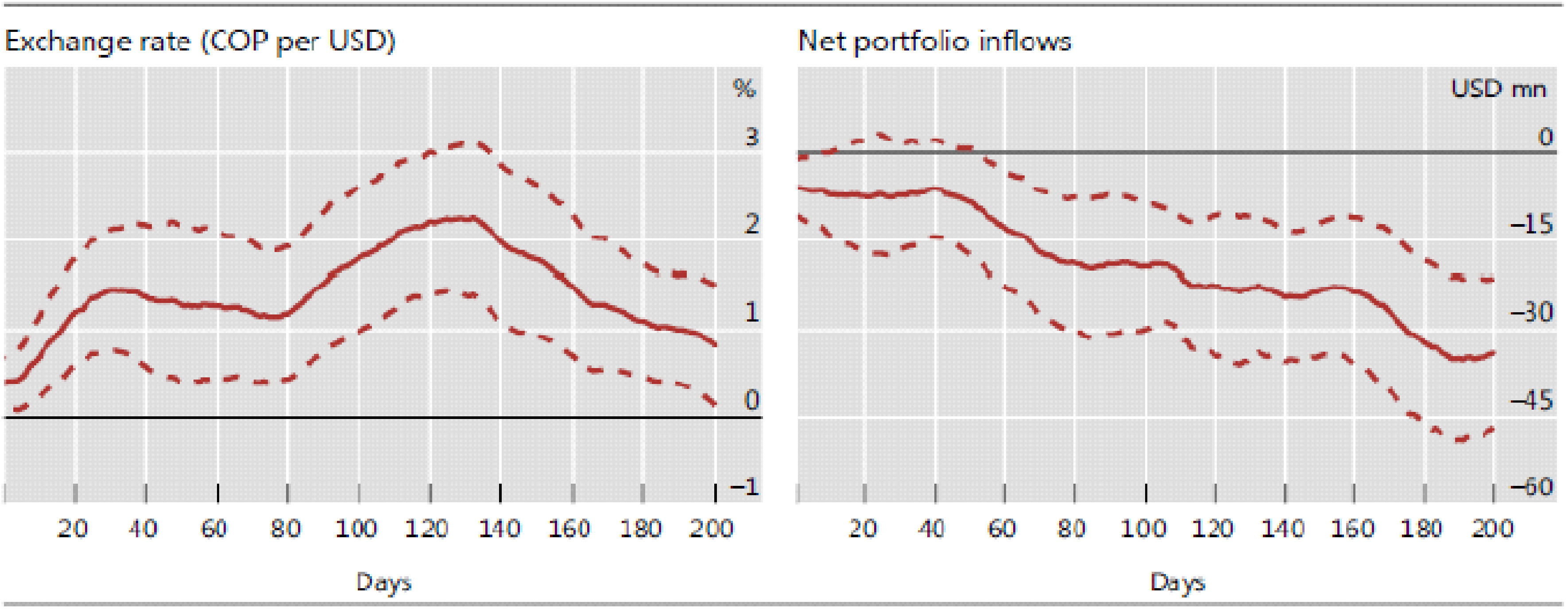
Impact of FXI on new corporate loans in Colombia



Size of impulse normalised to 30 million USD

Source: Hofmann, Shin, Villamizar-Villegas (2019)

Impact of FXI on exchange rate and capital flows in Colombia



Size of impulse normalised to 30 million USD

Source: Hofmann, Shin, Villamizar-Villegas (2019)

FX intervention impact depending on bank characteristics

	20 Days	40 Days	60 Days	80 Days	100 Days
Capitalisation	0.020***	0.039***	0.058***	0.077***	0.097***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)
Bank Size	0.014***	0.029***	0.044***	0.060***	0.077***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)
Debt	-0.003***	-0.006***	-0.01***	-0.013***	-0.016***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Provisions	-0.017***	-0.033***	-0.049***	-0.065***	-0.082***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
FXI*Capitalisation	0.10***	0.16***	0.19**	0.25**	0.31**
	(0.034)	(0.55)	(0.077)	(0.10)	(0.12)
FXI*Bank Size	0.11***	0.15***	0.18**	0.26***	0.31**
	(0.033)	(0.055)	(0.077)	(0.10)	(0.12)
FXI*Debt	-0.008***	-0.011**	-0.013*	-0.021**	-0.024**
	(0.003)	(0.005)	(0.007)	(0.009)	(0.012)
FXI*Provisions	-0.024	-0.036	-0.031	-0.035	-0.037
	(0.015)	(0.026)	(0.036)	(0.047)	(0.058)

Source: Hofmann, Shin, Villamizar-Villegas (2019)

References

Aldasoro, I, Borio, C and M Drehmann (2018): "Early warning indicators of banking crises: Expanding the family", *BIS Quarterly Review,* March 2018.

BIS (2018): "Moving forward with macroprudential frameworks", Annual Economic Report 2018

BIS (2019): "Monetary policy frameworks in EMEs: Inflation targeting, the exchange rate and financial stability", Annual Economic Report 2019

Gambacorta, L and A Murcia Pabon (2017): "The impact of macroprudential policies and their interaction with monetary policy: an empirical analysis using credit registry data", *BIS Working Papers* no 636

Hofmann, B and G Peersman (2017): "Is there a debt service channel of monetary transmission?", *BIS Quarterly Review,* December 2017

References (cont.)

Hofmann, B, Shin, H S and M Villamizar-Villegas (2019): “FX intervention and domestic credit: Evidence from high-frequency micro data”, *BIS Working Papers* no 774

Lang, J H, Peltonen, T and P Sarlin (2018): “A framework for early-warning modelling with an application to banks”, *ECB Working Paper* no 2182

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Measuring aggregate housing wealth: new insights from automated valuation models[1]

Susan McIntosh,

Board of Governors of the Federal Reserve System

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Measuring Aggregate Housing Wealth: New Insights from Automated Valuation Models

Susan Hume McIntosh
Flow of Funds Section
Board of Governors of the Federal Reserve System

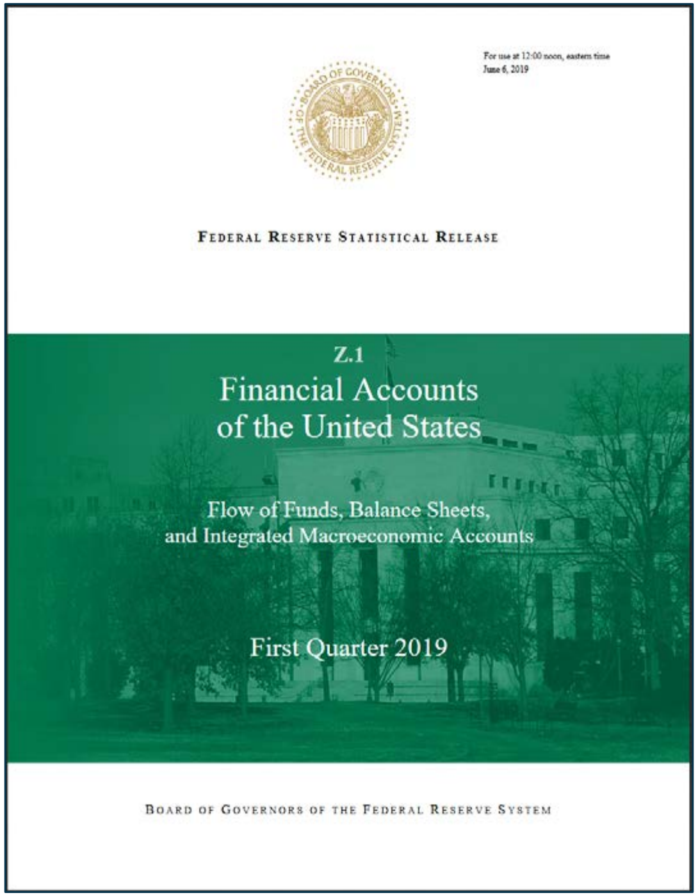
IFC-BNM-ECB Satellite Seminar on
"Post-crisis data landscape: micro data for the macro world"
Kuala Lumpur, Malaysia
August 16, 2019

Note: The analysis and conclusions set forth are those of the authors and do not indicate concurrence by the Board of Governors.

Why Care About Aggregate Household Housing Wealth?

- Affects many aspects of household financial decision-making: consumer spending, savings, small business formation, investment in education, geographic mobility
- Major component of total household wealth: value of the primary residence represents about 2/3 of a typical household's total assets (SCF 2016)
- Key driver of wealth changes during the Great Recession
- Frequent input into empirical and quantitative macro models and analyses

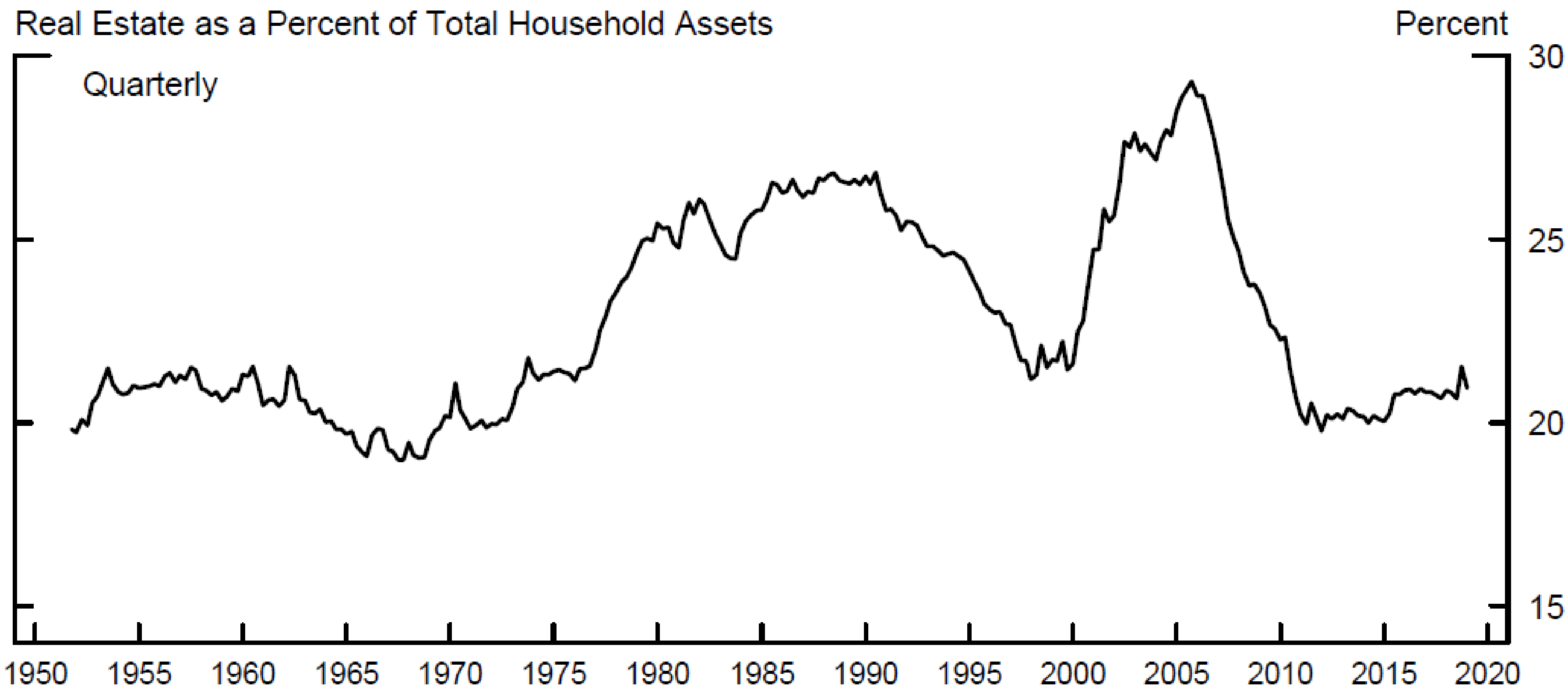
Financial Accounts of the United States



www.federalreserve.gov/releases/Z1/

- Time series:
  Annual 1945 – 2018
  Quarterly 1951:Q4 – 2019:Q1
- Includes flows, levels, some balance sheets, Integrated Macroeconomic Accounts, and more
- Published 10 weeks after the end of the reference quarter
- All data available in the Federal Reserve's Data Download Program (DDP)
  www.federalreserve.gov/datadownload/Choose.aspx?rel=Z.1

Household Residential Real Estate



Source: Financial Accounts of the United States, June 6, 2019.

Methods of Measuring Residential Real Estate Values

1. **Owners' reported values from nationally representative household surveys**: Survey of Consumer Finances (SCF), the American Community Survey (ACS), the biennial American Housing Survey (AHS)

2. **House price index**: currently use a repeat-sales housing index with an AHS benchmark to calculate housing value in the U.S. Financial Accounts (FA)

3. **New FRB method using an automated valuation model (AVM):** leverages machine learning and "big data" collected by Zillow

Owners’ Reported Values

Strengths

- Value entire stock of owner-occupied homes
- Captures quality changes

Weaknesses

- Over-optimism in normal times
- Lagged recognition of changing market conditions

Repeat Sales Housing Price Index

Strengths

- Market-price driven
- No reporting bias

Weaknesses

- Transacting homes only
- Holds quality constant
- Needs a benchmark value

Automated Valuation Model

Strengths

- Values entire stock
- Market-driven
- Captures quality changes, not just transaction pairs
- Model predictions testable against actual prices for traded units

Weaknesses

- Limited historical data
- Not nationally representative
- Modest model bias
- Rental properties included in AVM prices

Measuring the Market Value of Residential Real Estate

- **Current**: American Housing Survey (AHS) benchmarks (2005), fixed investment from BEA, and CoreLogic repeat-sales house price index

- **September 2019**: Zillow "big data" automated valuation models (AVM) for average prices and property counts from U. S. Census Bureau for quantities

What is Zillow?

(2006 , Seattle, former Microsoft employees)

Zillow Research Mission Statement:

"Zillow Research aims to be the most open, authoritative source for timely and accurate housing data and unbiased insight. Our goal is to empower consumers, industry professionals, policymakers and researchers to better understand the housing market."

Zillow Research Principles:

- Provides unbiased data and analysis about the housing market in a transparent way
- Is independent of Zillow's business goals and is not a revenue center
- Benchmarks findings against outside datasets to ensure accuracy and appropriate context
- Respects the integrity of data and uses it honestly

Why did we choose Zillow?

- Willing to provide data free of charge
- Very helpful in working with us
- Widely known in the U.S.
- One of the first companies to use this methodology on this scale
- Other vendors with related products did not offer clear advantages

Zillow's Data

What is Zillow's AVM?

- AVM source data from tax records when sales occur plus rich local data (school quality, water views, etc.)
- Thousands of models estimated frequently and combined to produce value estimates.
- Compares predictions against observed transactions

What did the FRB get?

- Average AVM estimates and counts of valued homes monthly, by county, for single family and multi-family separately, from 1996 to present
- Detailed information about coverage and model errors by geography, property type, and sales price

AVM Does Not Cover All Homes and Covers Some Rentals

Property Counts in 2017

(millions of units)

	ACS		Zillow
	Total	Own-Use	Total
Single-Family	92.9	76.4	78.7
Multi-Family	35.9	5.8	7.7
Total	128.8	82.1	86.3

AVM Value-Weighted Errors Are Close to 0

Source: Zillow and FRB calculations

Empirical Method

Price(AVM)* Quantity (ACS)

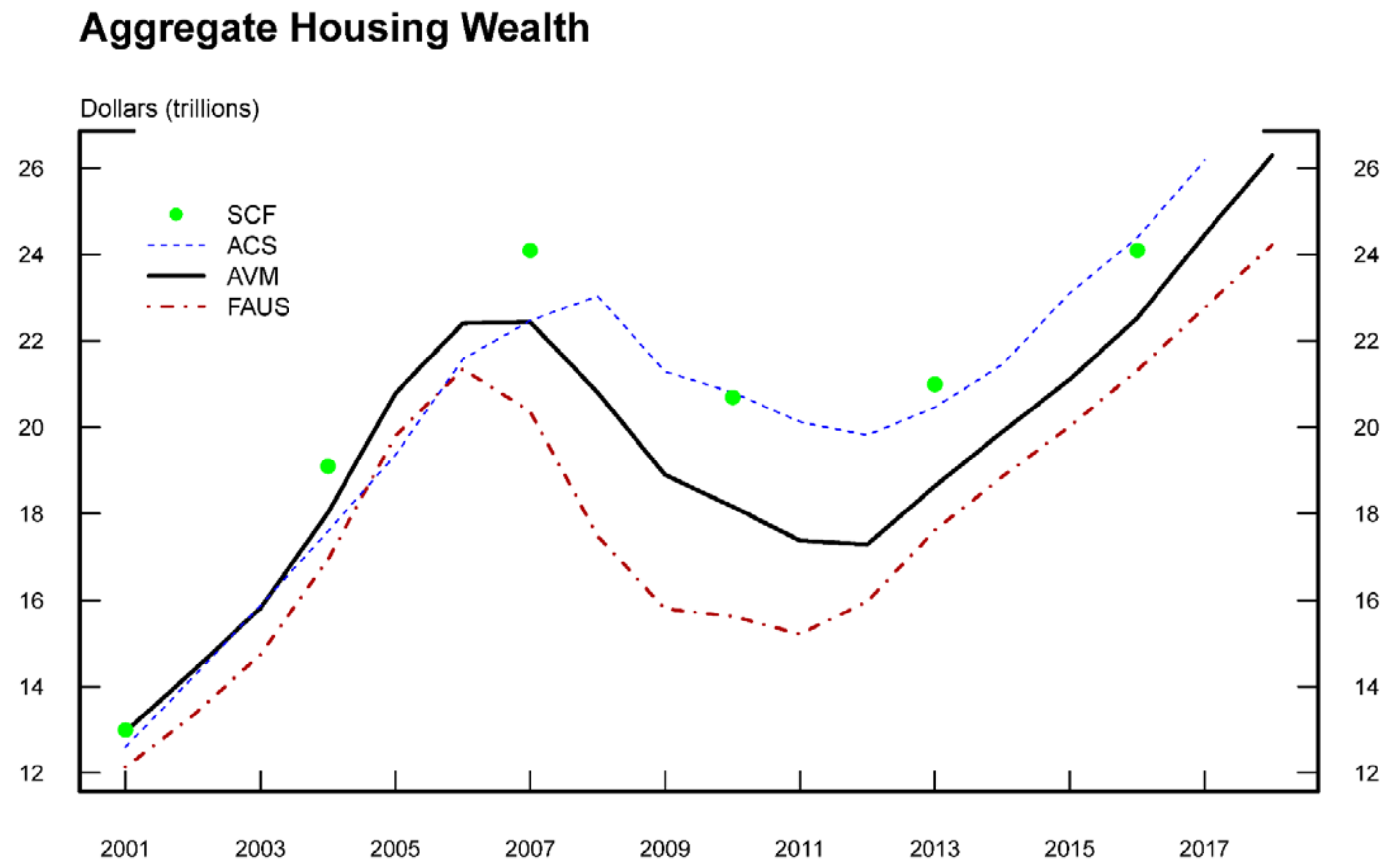
- **Problem:** Coverage issues prevent us from simply summing Zillow's AVM estimates, which are modestly biased

- **Solution:**
  - Use property count data from the ACS
  - For each county *i*, property type (mf or sf) *c,* and quarter *t*

$$\textit{Aggregate Value}_{i,c,t} = \textit{Property Counts ACS}_{i,c,t} \times \textit{Average value AVM}_{i,c,t}$$

  - *Average value AVM*$_{i,c,t}$ adjusted for model bias using value-weighted average error
  - Does not adjust for the inclusion of rental properties – likely a modest downward bias

Aggregate Housing Wealth



Source: American Community Survey (U.S. Census Bureau), Financial Accounts of the United States, Survey of Consumer Finance (triennial), and Zillow.

Conclusion

- AVM method overcomes known biases of prior wealth estimation methods
- AVM method "splits the difference" between owner-reports and House price indexes
- AVM measure is currently an Enhanced Financial Account (EFA) release:

https://www.federalreserve.gov/releases/efa/alternative-measure-of-owner-occupied-real-estate.htm

BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM

IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia and the European Central Bank

16 August 2019, Kuala Lumpur, Malaysia

Key indicators for sectoral productivity analysis[1]

Laurent Olislager,

European Commission

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

62nd ISI WORLD STATISTICS CONGRESS 2019

18 – 23 August 2019, Kuala Lumpur

Come | Connect | Create

IFC-BNM-ECB Satellite Seminar – Session 4
“Key indicators for sectoral productivity analysis”

Laurent Olislager

European Commission

16 August 2019, 15:45

Outline/Content

- Purchasing power parities (PPPs)
- PPPs for the production side of GDP
- Previous work
- Methods
- Data sources
- Results
- Sectoral productivity analysis
- Next steps

Purchasing power parities (PPPs)

- The purchasing power depends on the price of goods and services
- PPPs are indicators of price level differences across countries

- In Europe: Eurostat-OECD PPP Programme
- Worldwide: International Comparison Program

- Calculated for a wide range of goods and services → price level indices
- Used for real GDP comparisons, measurement of poverty rates, etc.

PPPs for the production side of GDP

- Usual method: collection and aggregation of consumer prices
- PPPs are thus calculated for the expenditure side of GDP

- But what about industry-specific productivity comparisons?
- Estimate PPPs for the production side of GDP: “PPPPs”

- Main issue: lack of reliable data sources
- Workaround: adjustments to existing sources

Previous work

- Methodology developed by University of Groningen (Netherlands)
- Used in several projects: EU KLEMS, WIOD, PWT

- Eurostat: similar approach, using only official European statistics
- First results: PPPPs for the year 2014 published in Eurostat EURONA

EUROSTAT REVIEW ON NATIONAL ACCOUNTS AND MACROECONOMIC INDICATORS
2/2016

Estimating purchasing power parities for the production side of GDP

LAURENT OLISLAGER AND PAUL KONIJN (¹)

Methods

- Adjusted expenditure PPPs + output PPPs
- Limitation: PPPs for the output (rather than value added) of industries

- Elteto-Koves-Szulc (EKS) calculation
- Basic headings (BHs): CPA/NACE 4-digit level

Data sources

- Only official European statistics
- Prices:
  - Prodcom: unit values for over 3000 manufactured goods
  - PPP Programme: consumer goods and services, equipment, construction, education, health, collective services
  - Agriculture prices and price indices: producer prices for 128 products
  - “Proxy PPPs”
- Weights:
  - Prodcom: production values at product level (below BH level)
  - Structural business statistics: GVA for 4-digit NACE branches (above BH level)
  - National accounts: GVA for 64 NA branches (further aggregation)

Data sources

A64*	NACE title	A64*	NACE title
1	Crop and animal production, hunting and related service activities	33	Air transport
2	Forestry and logging	34	Warehousing and support activities for transportation
3	Fishing and aquaculture	35	Postal and courier activities
4	Mining and quarrying	36	Accommodation; food and beverage service activities
5	Manufacture of food products, beverages and tobacco products	37	Publishing activities
6	Manufacture of textiles, w earing apparel and leather products	38	Motion picture, video and television programme production, sound recording
7	Manufacture of w ood and of products of w ood and cork, except furniture;	39	Telecommunications
8	Manufacture of paper and paper products	40	Computer programming, consultancy and related activities; information service
9	Printing and reproduction of recorded media	41	Financial service activities, except insurance and pension funding
10	Manufacture of coke and refined petroleum products	42	Insurance, reinsurance and pension funding, except compulsory social
11	Manufacture of chemicals and chemical products	43	Activities auxiliary to financial services and insurance activities
12	Manufacture of basic pharmaceutical products and pharmaceutical	44	Real estate activities
13	Manufacture of rubber and plastic products	45	Legal and accounting activities; activities of head offices; management
14	Manufacture of other non-metallic mineral products	46	Architecture and engineering activities; technical testing and analysis
15	Manufacture of basic metals	47	Scientific research and development
16	Manufacture of fabricated metal products, except machinery and equipment	48	Advertising and market research
17	Manufacture of computer, electronic and optical products	49	Other professional, scientific and technical activities; veterinary activities
18	Manufacture of electrical equipment	50	Rental and leasing activities
19	Manufacture of machinery and equipment n.e.c.	51	Employment activities
20	Manufacture of motor vehicles, trailers and semi-trailers	52	Travel agency, tour operator reservation service and related activities
21	Manufacture of other transport equipment	53	Security and investigation activities; services to buildings and landscape
22	Manufacture of furniture; other manufacturing	54	Public administration and defence; compulsory social security
23	Repair and installation of machinery and equipment	55	Education
24	Electricity, gas, steam and air conditioning supply	56	Human health activities
25	Water collection, treatment and supply	57	Social w ork activities
26	Sew erage; w aste collection, treatment and disposal activities; materials	58	Creative, arts and entertainment activities; libraries, archives, museums and
27	Construction	59	Sports activities and amusement and recreation activities
28	Wholesale and retail trade and repair of motor vehicles and motorcycles	60	Activities of membership organisations
29	Wholesale trade, except of motor vehicles and motorcycles	61	Repair of computers and personal and household goods
30	Retail trade, except of motor vehicles and motorcycles	62	Other personal service activities
31	Land transport and transport via pipelines	63	Activities of households as employers of domestic personnel and
32	Water transport	64	Activities of extraterritorial organisations and bodies

Results

- PPPPs for
  - 2008-2017
  - EU28, Switzerland, Iceland, Norway
  - 64 National Accounts branches (NACE A*64)

- Assessment: coverage, reliability, plausibility
- Example: comparison of price level indices at GDP/GVA level

Sectoral productivity analysis

- Comparison of productivity levels across countries
- At the heart of discussions about competitiveness

- PPP-adjusted labour productivity at A*64, A*21 and A*10 levels
- Limitation: single deflation rather than double deflation

Next steps

- Publish results as experimental statistics
- With detailed quality analyses

- Future work: address the biggest methodological challenges
- E.g. further experiment with data sources and adjustments

- Most importantly: extend coverage for intermediate consumption
- This would allow calculating value-added PPPs

62nd ISI World Statistics Congress

ISI 2019 Kuala Lumpur

18 - 23 August 2019 | www.isi2019.org

THANK YOU

Laurent.Olislager@ec.europa.eu

(thank you to all colleagues involved)