Focus on micro data:

Potential benefits for the industry?*

Professor Claudia M. Buch

Deutsche Bundesbank

8th European Central Bank Conference on Statistics

Central Banking Statistics. Moving beyond the aggregates

Frankfurt am Main, 6 July 2016

I. Focus on micro data: potential benefits for the industry?

This session addresses two issues that are closely interlinked. The first concerns the costs and benefits of using micro data in statistics in general and in central bank statistics in particular. The second deals with the benefits of micro data for the industry. These two questions cannot be answered independently of each other. Central bank statistics are a public good. Therefore, the costs and benefits of different types of statistics need to be assessed from the perspective of society as a whole.

The financial industry is a key stakeholder in this process. The quality of the data reported hinges on the reporting mechanisms employed by the industry, and data that are not regularly kept up to date by financial institutions may be of limited value for central banks, too. To a large extent, this aligns the interests of central banks and the reporting parties.

But, of course, public and private interests may differ when it comes to aspects such as reporting costs. While reporting costs need to be balanced against the benefits of statistics, some costs for the industry are inevitable in order to improve monitoring of the financial system and reduce the likelihood of costly financial crises occurring in the future.

In the following, the traditional statistical compilation process is compared with the new, microdata-based approach. This sheds light on the large potential of micro data, not only for the creation of analytical benefits, but also for cost-savings.

II. Central banking statistics from the traditional "value chain" to micro data

In the system of official statistics that has evolved over time, reporting forms are used to obtain information, which is used to compile specific statistics in a pre-defined aggregation form. The reports are collected, checked for plausibility, supplemented by estimates for missing data, extrapolated if need be, and then consolidated across pre-defined dimensions. It is often neither legally nor technically possible to use the underlying micro data later on. Firms whose reports have been included in these calculations can use the aggregated statistics to analyse the macroeconomic environment, for instance. However, focused studies on specific market segments are generally not feasible.

Micro data initiatives are based on a fundamentally different concept: data should only be collected once and be used to compile various statistics. This approach entails initial investment costs – for example, investment in information technology and training staff – as well as current expenditure for data collection and program maintenance. However, this information also has a much greater scope of use. Compared with traditional statistics, micro data offer a range of advantages for various user groups.

III. What are the benefits of using micro data?

Over the past decade, the availability and use of micro data have revolutionized empirical work on financial systems (Deutsch Bundesbank 2015a). Technological progress has contributed to improved access to micro data and to improved handling of large, granular datasets. Empirical methods that allow heterogeneity across suppliers and users of financial services to be explored have become readily available. Researchers both inside and outside policy institutions are highly skilled and use these improved infrastructures for their research and day-to-day analytical work.

This work has shown that aggregate developments cannot be fully understood without exploring the rich sources of heterogeneity shaping the response of market participants to shocks and regulations. Systemic risk, for example, is defined as the contribution of the distress of individual financial institutions (or a group of financial institutions) to overall stress in the financial system, with adverse repercussions on the real economy. The contribution of individual financial institutions to systemic risk is higher, the greater the risk of an individual institution, the larger an institution is (too big to fail), the more connected an institution is (too connected to fail), or the more financial institutions are exposed to common risk factors (too many to fail). This definition shows that systemic risk cannot be analyzed without making use of detailed and granular data on financial institutions (Deutsche Bundesbank 2015c).

Similarly, analysis of the transmission of monetary policy and of the effects of monetary policy on risk taking has been changed fundamentally by the availability of micro data and the parallel development of empirical models that allow these data to be analyzed (Deutsche Bundesbank 2015b). These methods enable drivers of demand and supply to be identified and the heterogeneous responses to monetary policy shocks of different types of financial institutions (small versus large, domestic versus foreign) to be studied. New methods permit a much more

detailed insight into the workings of financial and money markets than analyses using aggregate time series.

An improved understanding of (risks to) financial stability and of the transmission of monetary policy is a clear benefit that the use of micro data brings to central banks. This enables central banks to better perform their core tasks – maintaining price stability and contributing to financial stability. In this sense, micro data benefit both, society as a whole as well as financial institutions that operate in a more stable environment.

An additional advantage of an increased use of micro data is that these data can be used for multiple purposes. Traditional statistics tend to serve specific purposes, such as the construction of aggregate indicators for statistics on the financial account. If new analytical questions arise, these (aggregate) data often do not fit the purpose. Analysis then has to rely on information which is not really targeted, or additional surveys have to be conducted. In addition to being costly, these surveys often have the disadvantage that they do not allow developments to be tracked over time and that they survey only a sample of firms or banks and are thus not very representative. Using micro data that have already been collected, it is comparatively simple to respond to a new request by carrying out an appropriate aggregation of the data.

Looking ahead, the microdata approach to compiling statistics may even replace the traditional one, leading to cost-savings on the side of the reporting agents. In the traditional system, data has to be aggregated and reported separately for each dataset. The cost of aggregation is borne by the reporting agent. According to the microdata approach, reporting agents have to send their data only once. Everything else is done by the central bank. Hence, the costs of aggregation are passed on to the central bank.

That is a considerable advantage for the industry, but is that everything? Is it not the case that the industry not only benefits from cost savings but can also make direct use of the micro data? Can statistics be jointly compiled by the industry and the central bank on a voluntary basis in order to increase benefits and reduce costs?

This is not a purely hypothetical question. The Bundesbank's corporate financial statement statistics, which contains data on German non-financial corporations, are based on several sources, a key one being a data pool with financial statements provided on a voluntary basis by credit institutions and financial service providers. The individual institutions cannot access the

micro data of the other parties. However, they receive a very wide range of statistics on key ratios, broken down by sector and size category, which they can use as part of their credit assessment of firms or for sectoral analyses. For this reason, industry representatives have started to voice concerns about the collection of statistics which are insufficiently comprehensive and that cannot be used for detailed market analysis.

IV. What are the challenges and the costs of using micro data?

The above examples may suggest that the advent of micro data has moved us into a brave new world in which we have better information, better analyses, and a mutually beneficial exchange of information. But any cost-benefit analysis must, of course, also address the challenges and the costs that are associated with the use of micro data. Among those, data confidentiality and the costs of upgrading and maintaining IT systems feature prominently.

a) Data confidentiality and data protection

The traditional "aggregate" statistics that central banks use and produce generally do not raise data confidentiality issues. Whenever a sufficient number of observations (typically: three) have been aggregated, the behavior of individual firms, households and banks cannot be tracked from specific data points.

Micro data, by contrast, almost always raise confidentiality issues that may differ substantially across jurisdictions. For the United States, information about the profitability and the balance sheets of commercial banks has been available for many years in the "Call Reports" that are published by the Federal Reserve Bank of Chicago. These data have been used for many important studies on the US banking industry. It would not be possible to publish such data in many other jurisdictions. "Feedback loops" to the industry or other users – which would allow the behavior of individual firms or competitors to be identified – are not permitted. Perhaps even more importantly, micro data cannot be merged across jurisdictions, which complicates comparisons of analytical work for different countries.

In order to overcome this obstacle, new analytical and research networks have emerged that allow information and research results to be exchanged across countries without sharing the underlying (confidential) data. One example is the International Banking Research Network. Established in 2012, the IBRN brings together researchers from more than 20 central banks around the world – in addition to international organizations such as the BIS, the ECB and the IMF – in order to analyze current issues pertaining to international banking. The different country teams use comparable, country-level data on cross-border banking activities and common methodological tools to analyze issues such as the transmission of liquidity risk, spillovers of (macro)prudential policies, or the cross-border effects of monetary policy. As the studies use a common research methodology, the results are comparable across countries and can be analysed using meta-analyses (Buch and Goldberg 2015).

In each national jurisdiction, additional technical solutions are available that facilitate the use of micro data without compromising on data confidentiality. The Bundesbank, for example, has recently established a Research Data and Service Centre (RDSC), which provides researchers inside and outside the Bank with access to its micro data.

In general, the more detailed the data are, the more restrictive is the data access offered by the RDSC. Outside Bundesbank premises (off-site), only anonymised data specially prepared for academic research purposes (scientific use files) from the Panel on Household Finances study – a representative survey on the structure and composition of households' wealth – may be used.

Besides the Panel on Household Finances, the RDSC provides anonymised datasets on banks, securities, investment funds, enterprises and households, all of which can be accessed at dedicated researcher workstations inside the Bundesbank. For some datasets, remote execution is possible, too. External researchers have to send in their syntax files, which will be processed by the RDSC staff. For both of these means of access, the RDSC staff screen the output files, preserve anonymity, and send them to the external researcher.

The range of methods for access to data satisfies researchers' need for the rapid provision of high-quality data, while at the same time keeping sensitive information confidential. The Bundesbank is thus adhering to an established global standard: it is ensured that analytical findings that are to be published are anonymised such that no inferences may be drawn regarding particular individuals or enterprises. The international standard may be briefly summarised as follows: "Safe People, Safe Projects, Safe Settings, Safe Outputs, Safe Data".

b) Investment in (IT) infrastructure

Investment in IT infrastructure is a part of the construction of every new statistics (Schubert 2015). The procedures for processing micro data, which are only collected once, can be standardised to a greater extent than is the case for the individual silo procedures used to date. Accordingly, the multi-purpose approach will make it easier for both, data collecting authorities and reporting parties, to develop integrated systems. Two initiatives of the ESCB should be mentioned in this context:

1. The Banks' Integrated Reporting Dictionary (BIRD) is a document which helps to develop a standardised model for organising the banks' internal data warehouses in an integrated way, and for applying the transformation rules in the reports banks have to submit. The dictionary can be used on a voluntary basis. The advantages are a higher data quality, more efficient production of reports, more consistent data, a univocal interpretation, and clarity of regulations.

2. The European Reporting Framework (ERF) should be of great significance for the data collecting authorities in helping to ensure that data only have to be collected once, using a harmonised procedure for different purposes. The framework will cover most reporting requirements of the ECB and the European Banking Authority (EBA). The data reported in this framework will form a basis for harmonised production of secondary statistics.

V. How to make better use of micro data

The use of micro data in central bank statistics has many potential upsides, but providing and maintaining the necessary (IT) infrastructure also entails costs. As things stand, the optimal point of balancing costs and benefits has probably not yet been reached, and there are four key areas for improvement:

1. **Realigning micro data with existing statistics**: It will be possible to reap the full benefits of micro data in terms of efficiency and effectiveness of reporting only if there is close coordination between the scope of existing statistics and newly collected micro data. This

may, in some instances, require the scope of existing statistics to be adjusted, and it requires detailed planning when designing new data requirements.

2. **Improving the technical infrastructure**: Statistical reporting systems can be costly if they are not adequately integrated and linked to existing IT systems. Moreover, adjusting systems *ex post* is costly. The sooner potential uses for new statistics are known and the better reporting systems are aligned with existing IT infrastructures, the lower will be the costs of compliance and reporting.

3. **Improving feedback loops**: Setting up and improving feedback loops not only require close coordination and communication between central banks and commercial banks supplying statistical information. It also requires close cooperation within the private sector, involving potential suppliers of data and users of information provided in feedback loops. To compare the firms' own data to patterns in market data, firms' reporting entities (for example, the accounting department) and analysts need to exchange information. In particular, the quality of the statistics is likely to be enhanced if the reported results are analysed by the enterprise itself.

4. **Improving dialogue with the industry and academia**: Efficient designs of data requirements, IT systems, and feedback loops require good dialogue between central banks and data providers in the industry. At the same time, fostering the dialogue with academia can help improving analytical systems and exploit statistics to the best possible degree.

References

Buch, C.M., and L. Goldberg (2015), Mini-symposium on international banking, IMF Economic Review, Vol 63, No 3, pp 375-376.

Deutsche Bundesbank (2015a), Microdata – paradigm shift in central banks' statistics, Annual Report 2015, Frankfurt a.M., pp 47-59.
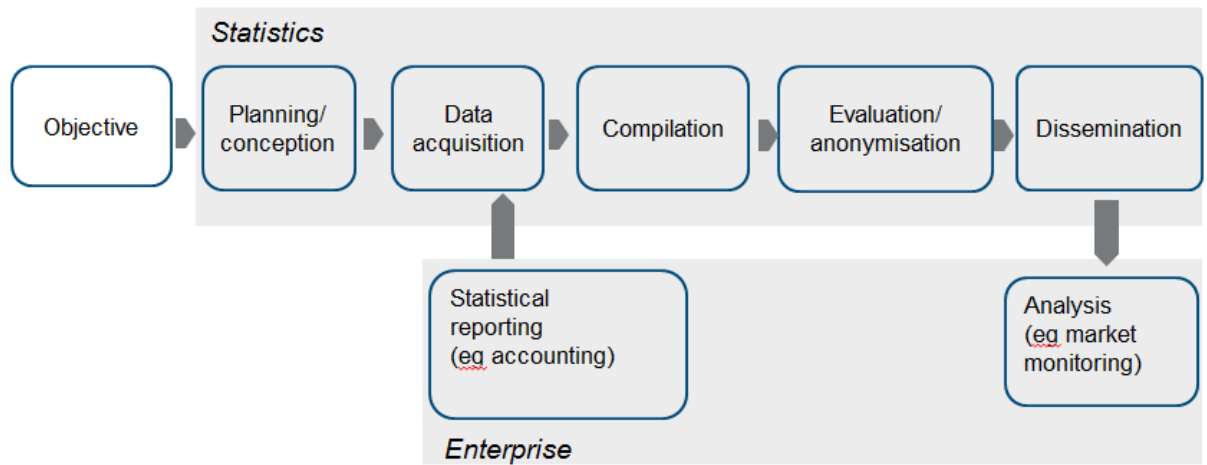
Deutsche Bundesbank (2015b), Effects of monetary policy on risk taking, Monthly Report, March 2015, pp 50-54.

Deutsche Bundesbank (2015c), Micro data and macroprudential policy, Financial Stability Review 2015, pp 80-81.

Schubert, A., (2015). Data as a core central banking asset – The strategy of the ECB, Presentation at the Big data workshop, Sveriges Riksbank, 9 September 2015.

The statistical value chain: traditional system

The statistical value chain: micro dataset (new)

Micro dataset

Diverse requirements → Uniform concept → Coherent data collection → Multi-dimensional compilation → Multifaceted evaluation and anonymisation → Separate dissemination, if applicable

Statistical reporting (eg accounting department) ↔ Analysis (eg market monitoring)

Enterprise