Motivation

João Cadete de Matos[1]

National central banks generate and have access to huge amounts of data. In spite of that, the issue of "data gaps" has been attracting a lot of attention within the central banks' community for some time now. Particularly in the wake of the financial crisis of 2007–09, the statistical systems have been experiencing additional pressure to expand their reach with a view to covering areas where the shortage of appropriate data had been more damaging – *e.g.*, information for the purpose of financial stability analysis and systemic risk assessment.

However, trying to keep up with the rapid changes of the economy and continuously adapting the statistics to new phenomena has some serious limitations. Conventional data collecting systems cannot simply keep on expanding indefinitely to cope with the need to fill the information gaps perceived by the users or in anticipation to their future data requirements.

In this context, it makes sense to exploit the largely unused statistical potential of the available micro-databases covering different areas of the economy and the financial markets – as in the case of, *inter alia*, records from existing central credit registers and business registers, and data related to the prudential supervision function. Once statistically edited, these micro-data might play an important role in enhancing the efficiency of central banks' statistical systems. The granular nature of this information, together with a good coverage of the relevant population, offers increased flexibility as regards the compilation of new statistics (*e.g.*, those related to financial and other structural innovations) and a more rapid response to *ad hoc* data requirements from the users. In general, this approach is technically easy to implement and with relatively low costs associated.

The evolution in network and communication protocols, database systems and multidimensional analytical systems has somewhat removed the potential disadvantages of having to deal with the huge amounts of data normally associated with the handling of micro-databases. In addition, these developments created the objective conditions for the statistical systems based on the so-called stove-pipe model, in which statistics in individual domains have developed independently from each other, to evolve to the next level, which is that of coherent and fully integrated data systems, enabling rapid data exploration, multidimensional analysis and cross-referencing of multiple sources with different granularities. With this objective in mind, many organizations have been developing, for the last two decades, their own data warehousing projects, ranging from combining multiple legacy systems to developing user interface tools for analysis and reporting.

This ongoing paradigm shift seems all together indispensable and irreversible, but is one that involves innumerable difficulties and challenges. One way to deal with these problems is to discuss them within the central banks' community and try to learn from each other experiences. The Porto Workshop on Integrated Management of Micro-databases is just one step in that direction.

---

[1]   Statistics Department, Banco de Portugal, jcmatos@bportugal.pt.

Structure of the workshop

In keeping with the Workshop's preliminary programme (see annex), the one and a half days of the event will be organised as follows:

1. On the first day, the whole morning and part of the afternoon will be dedicated to discuss the potential and actual uses of micro-databases for central bank statistics, in particular for macro-analysis. To accommodate as much as possible the high number of proposals for presentations on this topic that were received so far from central banks and international organizations around the world, this part of the Workshop will be organised in parallel sessions, devoted to describing the characteristics and uses of existing micro-databases, by statistical domain (*e.g.*, central credit registers, central balance sheet databases, security-by-security databases, balance of payments/international investment position data), as well as the possible uses for statistical purposes of other, less traditional, data depositories, such as data from payment and settlement systems and the *Bankscope* commercial database. In the ensuing panel discussion, the participants will attempt at drawing conclusions from the earlier discussions.

2. The second half of the afternoon and most of the morning of the following day will be centred on data integration issues, essentially from a front-end (*i.e.*, the users' side) standpoint. Participants will be encouraged to describe the way in which their organizations retrieve information from different sources (*i.e.*, from multiple micro-databases) and assemble it in a unified view, as well as the techniques they use and the problems they face in so doing. Discussions are also expected to tackle the methodological, statistical and technical issues associated with combining aggregate data with micro-data, the use of micro-data to generate macroeconomic and statistical information and the possibility to drill down through the data hierarchy to examine increasingly granular levels of detail associated with combining aggregate and survey data. The use of business intelligence technologies to help decision-makers use data, documents, knowledge and analytical models to identify and solve problems in a more effective way, will be addressed here as well.