Economic and Finance statistics

Block 4: Micro data for Statistics

Part 1 – Benefits, challenges & key aspects of financial micro data

Nicolas Griesshaber • Javier Huerga • Antonio Matas •
Arek Wyka (European Central Bank)

Spring semester 2021

Outline

Part 1: Introduction to financial micro statistics - Benefits, challenges and key aspects of financial micro data

1	Background
2	Financial micro data for statistics – Benefits and costs; Operation of micro databases
3	Data quality management
4	Key aspects for financial micro databases – International Standards; Identifiers and linking
5	Outlook

Part 2: Security-by-Security databases

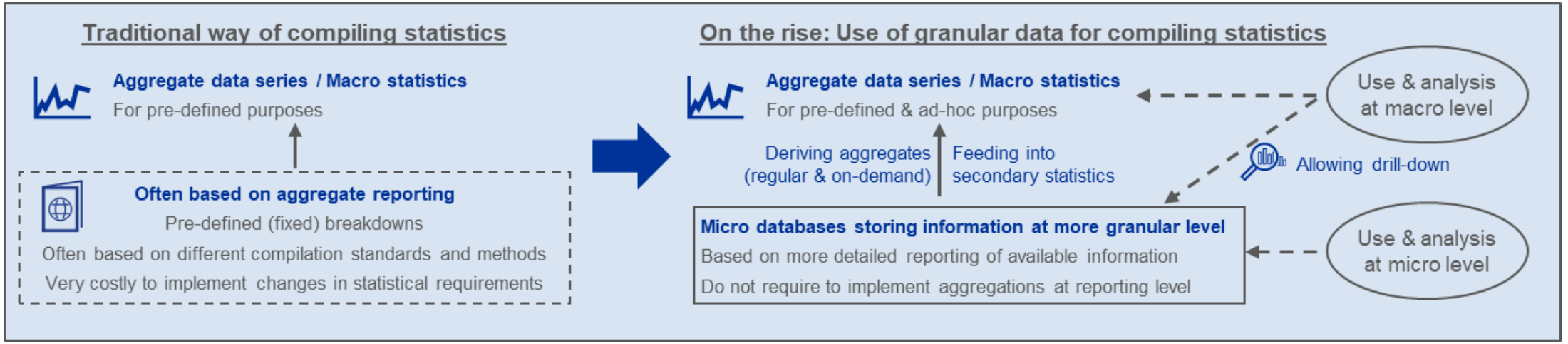
Part 3: Loan-by-loan data on credit and credit risk

Part 4: MMSR, €STR and Yield curve

Background

The rise of micro data for statistics

More granular data collections and micro-databases have become increasingly relevant for the production of statistics



Background

Why do we need micro data for statistics?

- Collection/compilation of aggregate financial, monetary and economic statistics has become increasingly burdensome
  - Difficult quality control of aggregated reporting
  - International statistical standards differ from market concepts used by reporting agents
  - Aggregate statistics are often not able to cope with financial innovation
  - Reporting burden increases with new or amended statistical requirements

➢ Need for a data collection system where reported information is simple and remains unchanged in case of new requirements, while data can be used in a flexible/agile way for multiple purposes

➢ **Solution: Granular data reporting** (e.g. security-by-security, loan-by-loan, etc.)

Financial micro data for statistics

Benefits and Costs

Advantages of micro data

- Increased harmonisation and data quality
  - Harmonisation in statistical concepts and calculation methods
  - Promote statistical standardisation, as discrepancies are easily recognised in the micro data
  - Quality checks can be done at very detailed level
  - Greater accuracy and consistency, as compilers rather than reporters are responsible for statistical classification of the data
- Easier implementation of statistics
  - Allows re-defining output aggregates without modifying the input data
  - Empowers compilers, while limiting reporting obligations for reporting agents → allowing increased timeliness of statistics

Financial micro data for statistics
Benefits and Costs

Advantages of micro data (continued)

- Flexible use and adaption
  - Flexibility to produce new aggregates on demand, going beyond totals and averages (e.g. dispersion)
  - Data can be flexibly combined to serve various requirements, while data quality needs to be checked only once
  - New statistical requirements usually do not require changes in reporting → Lower reporting burden
- Increased possibilities for analysis and use
  - Permits analysis of individual assets and institutions
  - Allows analysing the developments that are underlying the aggregates (drilling down aggregates)
  - Allows use of micro data in the design, implementation and monitoring of central bank or government policies

Financial micro data for statistics

Benefits and Costs

Micro-databases however involve some non-negligible costs

- IT infrastructure setting, maintenance and operation
  - Large data volumes
  - Granular data structure

  } Requires more sophisticated database infrastructure
- Data provision & revisions
- Data quality management

➢ **Technological advancements facilitate the use and operation of micro databases**

Financial micro data for statistics

Operation of micro databases

The operation of financial micro databases broadly consists of three stages

Data input

Data can be collected and/or purchased from a range of sources

- Central banks or government agencies
- Numbering agencies
- Commercial data providers
- Exchanges
- Regulatory or statistical reporting

Key aspects:

- Input standardisation (e.g. via data input dictionary)
- Definition of file format (xml, csv, json, etc,)
- Data source management (educate data providers)

Storage, DQM and compilation

Data receipt, merging and storage

Data quality management (DQM)

Checks for completeness, consistency and plausibility

- Internal consistency & stability over time
- Checks on individual & aggregate level
- Comparison with benchmark data

Data compilation

Enrichment of the data

- Calculations
- Default values

Production of a 'golden copy'

Data output

Micro data

Compilation of aggregates

- Not an easy task

Data dissemination:

Access to the data is a key issue

- IT tools
- Confidentiality aspects

➢ Different users may require different access restrictions or access routes (e.g. database access vs. output files)

Data quality management

Challenges and focus of data quality management for micro data

Complexity of data, number of data points and multipurpose use provide different challenges for data quality management (DQM) compared to macro statistics

- Standardised set of encompassing data quality checks (both at individual and aggregate level) that ensure sufficient confidence in the data
- Concise and transparent presentation of data quality to all parties (reporters, compilers, users)
- Ensuring 100% quality (no inconsistencies, no missing data, etc.) from all relevant user perspectives is hardly feasible (would require immense resources)
  - ➢ Important to ensure the most relevant potential data issues are detected and prioritised (requiring weighting of potential issues)

Data quality management

Comprehensive set of data quality checks

Checks can be usually distinguished in two types

- **Formal validity checks** (mostly automated)
  - Aim: Identify *errors with certainty*
  - Technical checks

    Ensure data is reported according to specifications
    - e.g. unexpected number of attributes
  - Business checks

    Ensure data is reported according to the business logic
    - e.g. mandatory input not provided

- **Plausibility checks**
  - Aim: Identify data that *may* not be correct (need to be validated)
    - Focus on consistency and stability over time;
    - Conducted both at individual and aggregate level
  - Internal checks → ensure data reported is consistent within database
    - Per reporting agent/data provider
      (e.g. consistency of data points; across time; outliers, etc.)
    - Across reporting agents/data providers
      (e.g. consistency of attributes on same type of information; outliers, etc.)
  - Comparisons with external benchmarks
    - Defined at aggregate level (e.g. comparing security-by-security data with balance sheet statistics; securities issues statistics)

➢ **Prioritisation is key for effective and efficient DQM**

Key aspects for financial micro databases
International standards

International standardisation is of key importance when defining micro data attributes

- In particular with regard to classifications and identifiers (e.g. country names/codes, instrument classifications, sector classifications, entity/instrument names, currencies, etc.)
- For many financial and economic data attributes, international statistical standards (e.g. SNA) or financial industry standards (e.g. ISO standards) can be relied upon
  - ➢ These should be applied whenever possible
    - o to achieve consistent and comparable statistics at a cross-country/global level
    - o to facilitate linking with other micro data sets
    - o even the structure of the files to transmit micro data can be an ISO standard (e.g. MMSR)

Key aspects for financial micro databases
International standards

Complementary use of financial industry standards
(e.g. ISO codes)

- Statistical standards are not always known to reporting agents
  - ➢ e.g. most market participants do not know the System of National Accounts (SNA), while they are familiar with the Classification of Financial Instruments (CFI) code
- Mapping of financial standards (e.g. CFI) to statistical standards (e.g. SNA) can support statistical compilation

Background information:

International Organization for Standardization (ISO)

- founded in 1946 (Geneva)
- 164 member countries (national standards bodies)
- 794 technical committees (TCs) and subcommittees in charge of standards development for different sectors
  e.g. TC 1 on Screw threads (est. 1947) – TC 324 on Sharing economy (est. 2019) ➔ TC 68 on Financial services (est. 1972)
- over 22,850 International Standards covering almost all aspects of technology and manufacturing

➢ ISO Standards …
- o respond to global market needs
- o are developed by global groups of experts (relevant industry, consumer associations, academia, NGOs and government) through a long and multi-stakeholder process
- o are based on consensus

(see also www.iso.org)

Key aspects for financial micro databases
Identifiers and linking

Financial micro databases often require the use of unique identifiers (e.g. for instruments, entities, etc.)

- To avoid duplications → example: unique identification of a security
- To simplify the reporting → instead of providing multiple classification details, just refer to the unique identifier
- To ensure quality → permits to receive and check data only once in a central manner
- To combine data from different data providers → e.g. on the same security or the same issuer/holder or borrower/lender
- To link the data internally → e.g. to group information of individual securities/loans under the issuing/borrowing or holding/lending entity
- To link the data with other databases → e.g. linking with entity reference information; linking securities, loan and derivatives data
- To exchange data externally → e.g. with other central banks or institutions

➢ Unique identifiers should ideally be **global and licence-free** to allow standardised use across databases of different economies → e.g. Legal Entity Identifier (LEI) for institutions; International Securities identification Number (ISIN) for Instruments

Key aspects for financial micro databases
Identifiers and linking

Let's say that reporting agents have to provide info on the counterparties of their operations

- **Without a unique identifier**
  - Each reporting agent needs to …
    - ✓ independently find out a number of details for each counterparty (e.g. economic sector, country)
    - ✓ Repeat all the details of a counterparty or an instrument every time it is reported
  - It is not possible for the statistical compiler to …
    - ✗ check correctness of the data
    - ✗ put together transactions for the same counterparty
    - ✗ link to other sources

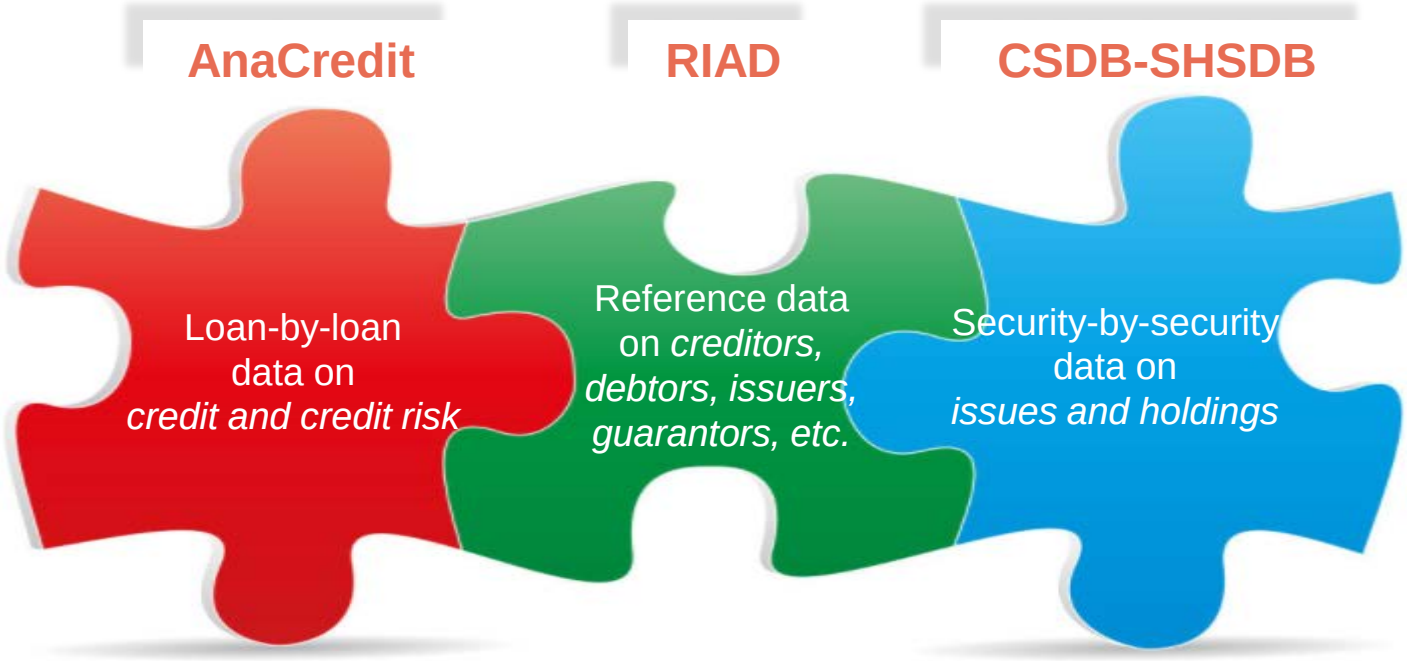
- **With a unique identifier**
  - Just the identifier of the counterparty is reported (always the same)
  - The statistical compiler uses the identifier to retrieve the necessary information from a reference database
    - ✓ Ensures higher quality and multiple possibilities of checking, grouping information, etc

Key aspects for financial micro databases

ECB's broader strategy for statistics

The availability of granular data has been gradually increasing at the ECB

- A holistic approach is followed with a view to allowing information from different frameworks to be combined
  - ➢ Repository of counterparty data plays a central role in this regard



Outlook

The remainder of this lecture aims to further address the aspects mentioned, focusing on some concrete examples of relevant financial micro data collections and databases

- Security-by-security (SBS) databases
- Loan-by-loan data on credit and credit risk (AnaCredit)
- MMSR, €STR and Yield curve

Thank you

Happy to receive any question or comments at the Q&A session on 13 July, 14 c.t.