



RESEARCH DATA CENTER

accesso ai microdati per la ricerca economica

How to facilitate output release



BANCA D'ITALIA

EUROSISTEMA

November 2024

When working on the microdata made available by Banca d'Italia, you are required to ensure that the output of your analysis complies with confidentiality standards, i.e. that it is not possible to trace the basic unit to which the information relates. Confidentiality requirements apply to any and all of your processing operations; failure to comply will make it impossible to release the processing output to you.

Any attempts to breach confidentiality may lead to your user account being suspended or removed.

Please find below a non-exhaustive list of confidentiality rules and guidelines to comply with, which will facilitate output release. The final decision on whether the processing output can be released lies with the Research Data Center, which assesses compliance with confidentiality requirements based on a complex set of variables.

Confidentiality rules

Please find below a not-to-do list for the purposes of confidentiality requirements and output release:

- Refrain from any calculation that may produce a single statistical unit as a result, including printing a single record or parts thereof or trying to export data;
- Do not screen print any statistics based on a single statistical unit, including minimum, maximum, median and quantile statistics;
- Do not create frequency distributions on continuous variables;
- Do not print any kind of chart;
- Do not screen print any residuals or forecast values or calculate their order statistics;
- Do not use any commands to connect external sources or change the work directory;
- Do not produce aggregate statistics on less than 3 statistical units; for this purpose, the output must show the number of observations underlying the calculations.

General guidelines

The following guidelines are intended to make reading the output simpler, thus facilitating its review and release:

- The code file produced must be clear, annotated and easy-to-read (including in terms of indentation), and the variables must have self-describing and unequivocal names;
- The number of jobs submitted and the expected output volume must be limited to what is strictly required for the research project. Any output exceeding 2,500 lines will be rejected;
- The number of observations included in the statistical models must be well higher than the number of parameters;
- For statistical models with very long outputs or a large number of parameters (e.g. many dummies or fixed effects) it is advisable to use commands for output abbreviation (e.g. the 'quietly' option in Stata).

The above list of criteria applies to each dataset and may be subject to changes and additions where required. Additional rules may apply to specific datasets.

Additional rules for specific datasets

Ancillary tables (BANCI)

For output based on banking group data, the reference statistical unit is the group (idba_capogruppo_ps variable). Therefore, aggregate statistics must cover at least 3 groups.