



RESEARCH DATA CENTER

REX - Remote Execution / Sistema di elaborazione a distanza

Instructions for the release of the outputs of the REX remote execution system



December 2022

Researchers working with the microdata made available by the Bank of Italy via its remote processing system shall ensure that the outputs of their analyses meet the confidentiality requirements, i.e. that they cannot be traced back to the respondent to which the information refers. Failure to comply with the confidentiality requirements, not only for single processing operations but also for a set of operations performed by the same user, shall result in the impossibility to provide the researcher with the outputs. Attempting to breach the terms of confidentiality may result in the credentials being suspended or revoked.

Below is a non-comprehensive list of rules and guidelines to help you comply with the confidentiality requirements and obtain the outputs. The final decision on whether the output may be released for publication lies with the Research Data Center, which shall verify compliance with the confidentiality requirements based on a complex set of variables.

How to comply with the confidentiality requirements

To avoid violating any confidentiality rules and, consequently, being denied access to the outputs, when writing your processing programs please make sure to avoid the following:

- any calculations that may give as a result a single statistical unit, including commands to print a single record or parts of the record, or to try to export the data;
- printing statistics calculated on a single statistical unit, including minimum, maximum, median and quantiles;
- creating frequency distributions on continuous variables;
- printing of any type of chart;
- printing residuals and predicted values and calculating their order statistics;
- commands to connect external sources or change the work directory.

General guidelines

The following principles aim to make it easier to read the outputs to facilitate the output control process (thus facilitating the release of the processing outputs):

- the code submitted must be clear, commented, easy to read (correctly indented) and the variables must be defined with self-descriptive and unambiguous names;
- the number of jobs submitted and the volume of expected outcomes must be limited to what is indispensable for the research; very long outputs (over 2,500 lines) will be rejected;
- the number of observations included in the statistical models must be adequately greater than the number of parameters;
- for models with very long outputs or with a large number of parameters (e.g. with many dummies or fixed effects), the use of commands for output abbreviations is preferable (e.g. the 'quietly' option in Stata)
- aggregate statistics must not be calculated on a small number of statistical units (less than three) and the output must always show the number of observations underlying the calculations.

The criteria listed above apply to each dataset and can be modified and supplemented whenever necessary. Additional restrictions may apply to some specific datasets.