EUROPEAN CENTRAL BANK

EUROSYSTEM

ANNEX 3

CRITERIA FOR OUTPUT CONTROL

1. Introduction and legal basis

This document sets out a series of rules, recommendations, and best practices to ensure that the work carried out by external researchers using Eurosystem microdata is disseminated securely. They are based Art 8(1) and 8(3) of Council Regulation 2533/99 and in particular on Art 2(3) Guideline ECB/1998/NP28[1]. If these requirements are not fulfilled, output cannot be released and/or publication cannot be approved. In the event of particularly serious breaches, additional measures may be taken.

2. Output Control Principles

The following rules aim to facilitate compliance with data confidentiality regulations.

1. Non-extraction of identifiers: Identifiers cannot be included in the results to be extracted or in the codes that generate them.

2. Non-extraction of microdata: No result may contain microdata. This entails refraining from extracting subsets of data, as well as tables, graphs, codes, or log files that contain microdata themselves. Consequently, the extraction of minimum and maximum values is not permitted.

3. Minimum number of observation units: All the results to be extracted should be based on at least three different economic agents. This applies both to aggregate results (averages, medians, etc.) and to charts and tables (at least three economic agents per cell/information node).

4. Degrees of freedom: Regression models must be calculated with at least ten observations and must also have at least ten degrees of freedom.

5. Dominance Rule: It is necessary to ensure that the largest economic agent does not exceed 85% of the total weight of the analysed value or any other weighting used.

Example: For calculating total sales in a specific sector for a particular year, let's consider only 3 banks. The total volume is 100 million euros, with the following composition: 90 million from the

[1] Art 2(3): "*all published data to cover at least three economic agents, and that when one or two make up a sufficiently large proportion of any observation, the data shall be arranged in such a way as to prevent their indirect identification*".

largest bank, 6 million from the second largest bank, and 4 million from the smallest one. In this case, the largest bank is potentially identifiable due to its contribution to the total value.

6. Confidentiality in multiple tables, control of differences: If the results are calculated based on a G population, but are subsequently recalculated for an X subset of G, the rules explained above must be met for observations of the difference. Otherwise, the individual observations could be identified on the basis of the differentiation.

Example: we have a table with all the banks in a given year and another with the banks in that year that exceed an X volume of deposits. We would have to create a third table with the banks that do not reach such X volume and check that the confidentiality criteria are met in that table; otherwise, the banks could be identified by differentiation.

7. Dichotomous (0-1) Categorical Variables (Dummies): When calculating averages of these variables, there should be a minimum of three economic agents for each category (three economic agents with 0 and three with 1).

8. Treatment of zeros and missing values: Zeros are permitted in regressions and descriptive statistical analysis, provided they do not represent missing values in dichotomous and categorical variables. In descriptive statistics, missing values will not be taken into account for determining the number of different economic agents used. If missing values are imputed, the number of imputed and observed values should be reported.

9. Additional criteria: The ECB or the DAP responsible for disclosure control may, insofar as it is deemed necessary, check the calculation results against additional criteria.

10. Before researchers can use calculation results outside the data access points (DAP), the supervising DAP must first check whether the researchers have applied the output control principles to them, and the calculation results must be approved for release by supervising DAP. The release of calculation results shall require the prior written consent of the supervising DAP.

3. Publication Control Principles

The following guidelines aim to assist researchers in more easily complying with the publication control standards ('publication control').

1. Review of all publications: Publications may not be published until the supervising DAP has given written approval. Approval may be withheld if the results intended for publication do not comply with the criteria below.

2. Reference to previously approved calculation results: Publications may only contain approved calculation results.

3. Referencing sources: The researcher commits to mentioning the ultimate data source in any publication resulting from this study, as indicated in the respective guide of each database.

4. Referencing charts and tables: All charts and tables should be referenced as follows: “Source: ECB, <name of the set of microdata used>, <period during which the microdata were used>, own calculations.”

5. Specification of type of data access: Each publication must specify the type of access the researcher had to the data, e.g. remote execution, on-site access (location) remote access or a combination of these access modes.

6. Copy of Publications: It is the responsibility of researchers to provide a copy of the published works that they produce and that contain research results from the analyses conducted on the accessed datasets.