

**UNITED NATIONS STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE  
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION  
STATISTICAL OFFICE OF THE  
EUROPEAN COMMUNITIES (EUROSTAT)**

**Joint ECE/Eurostat work session on statistical data confidentiality**  
(Luxembourg, 7-9 April 2003)

Topic (vi): Software tools for statistical disclosure control

**CELL SUPPRESSION IN EUROSTAT ON STRUCTURAL BUSINESS STATISTICS –  
AN EXAMPLE OF STATISTICAL DISCLOSURE CONTROL ON TABULAR DATA**

**Invited paper**

Submitted by Eurostat<sup>1</sup>

---

<sup>1</sup> Prepared by Paul Feuvrier (paul.feuvrier@cec.eu.int).

## Cell suppression in Eurostat on Structural Business Statistics - an example of Statistical Disclosure Control on tabular data

Paul Feuvrier  
Eurostat Unit D2 Structural Business Statistics  
Tel: (+352) 430 133 881  
Paul.Feuvrier@cec.eu.int

### 1 Confidentiality rules relevant for Structural Business Statistics

Structural Business Statistics consist of magnitude data (turnover value added, employment) combined with frequency data (number of enterprises, number of local units).

Statistical disclosure on SBS has been mainly controlled through cell suppression so far. That is Eurostat D2 has a very limited experience of perturbation-based methods, yet quite a significant one on restriction-based methods.

As far as primary disclosure is concerned, most countries use the (n,k) rule, which can be described as follows:

*"Regardless of the number of respondents in a cell, if a small number (n or fewer) of these respondents contribute to a large percentage (k percent or more) of the total cell value, the so-called n respondent, k percent rule of cell dominance defines this cell as sensitive."*

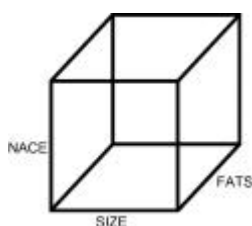
Yet some countries start to use the p% rule, where:

*"A table cell is declared sensitive, if upper and lower estimates for the respondent's value are closer to the reported value than a given percentage p."*

In the framework of the p% rule, one considers that the intruder is either the second largest contributor or more broadly a coalition of respondents pooling their data in an attempt to estimate the largest reported value.

### 2 Overlapping tables

All SBS related data are broken down by activity, ie by NACE Rev.1. In the SBS size class series, data are broken down by activity and size class. In the FATS series, data are broken down by activity and nationality of ownership. The example below shows three dimensions.



The three tables (activity, activity \* size class, activity \* nationality of ownership, ie FATS) are intersecting through the NACE. Thus, the confidentiality treatment needs to be co-ordinated between the different tables. Otherwise one figure might be hidden in one table and not in another - and the protection would be broken. This operation, where confidentiality flags have to be synchronised between the tables on every intersecting dimension, is called table to table protection.

### 3 SBS experience with GH-MITER via CIF

Eurostat Unit D2 has been testing and using for statistical production GH-MITER through CIF (Confidentiality Interface) since Autumn 2001. Indeed, the German engine met Eurostat requirements both in terms of treatment of dominance and table to table protection. GH-MITER uses the rectangular

parallelepiped method to smoothly protect data against absolute and approximate disclosure. It suits particularly well large tables or tables with more than two critical dimensions. A protection of these tables by hand would be either impossible or very time consuming.

Yet the "hyper cube criterion" is a sufficient but not necessary criterion for a safe suppression pattern. While testing and using the software, Eurostat pointed out significant over protection patterns. It seems that, for particular sub tables it has worked on, the best suppression pattern may not be a set of hyper cubes, in which case the hyper cube method would miss the best solution.

#### **4 Tau-Argus about to be tested on SBS data**

Indeed a new disclosure package Tau-Argus will be tested in Eurostat in April 2003. The paper will highlight these planned tests.

Indeed, Eurostat relies on this new system to produce less over protective data sets, while still smoothly implementing the confidentiality rules. The strong point of Tau-Argus is the calculation of an optimal set, where the hyper cube method is confronted to other techniques, e.g. the Hitas method, suppressing cells in hierarchical tables just like GH-MITER.