Working Paper No. 22 (Summary)

ENGLISH ONLY

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

USING DIS TO MODIFY THE CLASSIFICATION OF SPECIAL UNIQUES

Invited paper

Submitted by the University of Manchester, United Kingdom[1]

[1] Prepared by Mark Elliot (mark.elliot@man.ac.uk) and Anna Manning.

Using DIS to modify the classification of Special uniques - Abstract

Mark Elliot, Confidentiality and Privacy group, CCSR
Anna Manning, Centre for Novel Computing
University of Manchester

The special uniques detection algorithm (SUDA) has been shown to be effective in identifying risky records; Elliot (2002). By using High Performance Computing it is now possible to provide a comprehensive cross file risk assessment much ameliorating the problem of key variable selection in practical disclosure limitation; Elliot et. al. (2002). The method is now being used in the UK for the 2001 census outputs and labour force survey.

One issue with the current methodology is that the output metrics it produces are *ad hoc* with respect to the given key variable cross-classification and therefore the interpretation of them is an onerous and inexact process. This paper describes a new method that uses the Data Intruder Simulation (DIS) algorithm, (which provides an accurate file level measure of disclosure risk; Elliot 2000, Skinner and Elliot 2002), to calibrate the outputs of SUDA, thus deriving a per-record measure of the risk of correct matching which combines the intruder perspective value of DIS with per record nature of SUDA.

A pilot empirical study is described indicating a good relationship between the underlying correct matching probability (measured by the reciprocal of the population partition class) and the probability estimated by the combined DIS-SUDA method. Other alternatives for the combination method are discussed and necessary future work is outlined.