**Improving Disclosure Control for BRES**

**James Scruton (Office for National Statistics)**

**Summary**

A new product has been developed by UWE to apply disclosure control to the Business Register and Employment Survey (BRES). The protection tool has the potential to provide significant savings to the ONS in terms of time spent manually applying secondary suppression to tables. It also has particular benefits for ONS's Data Explorer tool, which aims to provide a flexible approach to data dissemination.

The tool has been completed and is in the testing phase. A decision on whether the tool will be accepted or not will depend on the outcome of this testing.

**Introduction**

The system used to process the results for the Business Register and Employment Survey (BRES) performs primary suppression of the BRES published tables. However, there is no automatic process for secondary suppressing the tables. The disclosure tool Tau Argus was investigated but found to be not appropriate because a) It did not fully protect the tables and b) cells were selected for suppression that were unsuitable. Therefore, secondary suppression is carried out manually by ONS, which is both time consuming (taking approximately two months), and error prone. The BRES team have been working with an academic from the University of the West of England, Dr. Martin Serpell, to try and automate the process of secondary suppression. He has developed a tool which has the potential to perform secondary suppression on large multi-dimensional BRES tables. This would allow any possible breakdown of the results to be produced, with no risk of disclosure. Preliminary work on this had shown a great deal of promise, and a Quality Improvement Fund (QIF) bid was submitted to secure the services of Dr. Martin Serpell to produce a tool that could be used to carry oit secondary suppression on BRES. The tool has generated a great deal of interest in both Methodology directorate and other business areas such as the Annual Business Survey (ABS).

**Background**

There are a number of different methods of disclosure control used in ONS. These often depend on the type of data in question i.e. rules to be applied to social survey data may differ from those used in business surveys. For example, for LFS/APS a simple threshold rule is applied with secondary suppression just taken care of by rounding on output. For most business surveys, however, the rules are often more strict.

The standard primary disclosure rules for BRES are:

1) Threshold rule - at least 3 enterprises in a cell

2) p% rule - The total of the cell less the 2 largest local units must be greater than or equal to a specified percentage of the value of the largest local unit.

The threshold rule is put in place to ensure that any person or business, not a member of the cell, can not identify a cell respondent or deduce the exact value or a close approximation of the response. The p% rule is put in place to ensure that a business contributing to a cell can not identify another business contributing to the cell or deduce an exact value or a close approximation of the other's response.

These rules produce a large number of primary suppressed cells at the level of detail required by users, which then results in a large number of secondary cells (up to three for each primary cell).

Most protection tools available only protect 2-dimensional tables (Tau Argus can protect 3-dimensional tables, but is unable to successfully handle datasets such as BRES), and suffer from the following problems:

* They are limited to small tables
* Only protecting 2-dimensional tables can lead to disclosure through linking tables
* Information not published is still vulnerable and can be deduced from information within the 2-demnsional tables

The UWE tool uses a hybrid algorithm that combines a Linear Program with a Genetic Algorithm to protect hierarchical multi-dimensional tables containing over 100,000 cells. The UWE tool has been demonstrated protecting a non-hierarchical 3-dimensional table with 1,000,000 cells. However it is its ability to protect large hierarchical tables that is of interest to ONS.

The tool will offer the following benefits to the BRES results team within ONS:

* Tables with 3+ dimensions can be fully protected
* Any table within those dimensions can be easily produced and published while being fully protected.
* The tool will save resource as it replaces the current process of manually applying secondary suppression

**Outline of work carried out so far**

The project plan included the following work items:

* Create a programme to manage the protection of BRES data
* Create a tool to optimally partition the BRES datasets prior to protection
* Create a tool to recombine datasets and remove excess secondary suppression
* Generalise above tools to work with other datasets
* Test tool

All apart from the last bullet point were completed before the project deadline (31 March 2014). The testing will largely be carried out by ONS, although UWE will be available to assist with this where necessary.

**Outline of work still to be carried out**

The project is in its testing phase. The testing will involve comparing the results from UWE’s protection tool against the current method of manually secondary suppressing the BRES information.

It has already been shown that the tool offers complete protection as oppose to the current method which can be unpicked using sophisticated methods. However, acceptance of the tool will also depend on the following criteria:

* Amount of data suppressed by tool
* Quality of suppression
* Time taken to produce a good quality
* Robustness of the tool
* Support within ONS (methodology and Information Management)

**Further uses of the protection tool**

The tool has the potential to dramatically improve the way that disclosure control is carried out across the whole of the GSS. Any dataset could in theory be suppressed in such a way that any cut of the data can subsequently be made with no risk of disclosure. This has particular benefits for ONS's Data Explorer tool, which aims to provide a flexible approach to data dissemination. However, other surveys, particularly those dealing with large, hierarchical information could see big benefits by using the tool.