CONFID2: STATISTICS CANADA'S NEW TABULAR DATA CONFIDENTIALITY SOFTWARE

Olena Frolova, Jean-Marc Fillion, Jean-Louis Tambay[1]

ABSTRACT

Under the Statistics Act, Statistics Canada is obliged to protect the confidentiality of its respondents' data. To meet this obligation, the cell-suppression technique is often used to protect tables of magnitude data. The automated confidentiality software CONFID, developed at Statistics Canada in the 1980s, is presently used for that purpose. However, user needs have continued to grow. To meet those needs, Statistics Canada has recently developed a new SAS-based confidentiality program: CONFID2. The program applies a suppression pattern to large tables that is not constrained by their number of dimensions. This paper covers the methodology, functionality and characteristics of CONFID2. It also provides some indication of the software's performance and of preliminary user feedback.

KEY WORDS: Confidentiality, Disclosure control, Objective function, Residual suppression, Sensitivity

RÉSUMÉ

En vertu de la Loi sur la statistique, Statistique Canada est tenu à protéger la confidentialité des données de ses répondants. Pour ce faire, on recourt souvent à la technique de suppression de cases, qui permet de protéger les données tabulaires quantitatives. À l'heure actuelle, on se sert du logiciel de confidentialité automatisé CONFID, développé à Statistique Canada dans les années 1980, pour faire cette opération. Toutefois, les besoins en la matière n'ont cessé de croître. Afin d'y répondre, Statistique Canada a récemment développé un nouveau logiciel de confidentialité fondé sur SAS: CONFID2. Ce dernier applique un modèle de suppression aux grands tableaux indépendamment de leur nombre de dimensions. Cette présentation couvre la méthodologie, la fonctionnalité et les caractéristiques de CONFID2. De plus, elle fait état de la performance du logiciel ainsi que de la rétroaction préliminaire de ses utilisateurs.

MOTS CLÉS: Confidentialité; contrôle de la divulgation; fonction objective; sensitivité; suppression résiduelle.

1. INTRODUCTION

1.1 Description of the problem

CONFID2 [1] is an improved version of Statistics Canada's (StatCan) tabular data confidentiality software, CONFID. CONFID [2] uses the cell-suppression technique to protect confidential cells in tables. Developed in the early 1980s, the system has received a number of enhancements over time but users wanted a version that was more user-friendly and that had more capabilities. In terms of user-friendliness, users would like to have a SAS program with more flexibility for data manipulation using standard SAS tools. They prefer to have a program that could be integrated into a larger SAS program, and that is open to integrate new features in the future (for example, processing negative values). It was also highly desirable for the new system to be created in the same manner as other generalized systems at StatCan, and for it to make use of existing LP solvers such as the SAS/OR® LP solver. In terms of improved capabilities, the new software should be able to create a suppression pattern for n-dimensional problems, where n is limited only by hardware limitations. In addition, the program should be much more time efficient, should be able to handle alpha-numeric codes and process large tables with more than 50,000 cells. Moreover, the software should be able to process microdata by groups and handle multiple decompositions. To satisfy all these needs the new confidentiality software CONFID2 was developed.

1.2 Organization of the paper

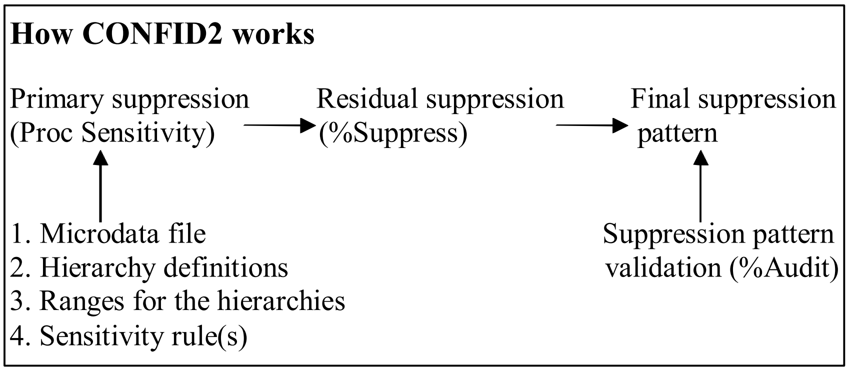
In section 2 we will discuss the methodology of CONFID2, in particular the methodology of the procedure SENSITIVITY and the macros SUPPRESS and AUDIT. In section 3 we will compare results produced by CONFID and CONFID2. Feedback from users and discussion of future work, presented in section 4, will conclude the paper.

Olena Frolova, Statistics Canada 100 Tunney's Pasture Driveway Ottawa, Canada, ON K1A 0T6, Olena.Frolova@statcan.gc.ca
Jean-Marc Fillion, Statistics Canada 100 Tunney's Pasture Driveway Ottawa, Canada, ON K1A 0T6, Jean-Marc.Fillion@statcan.gc.ca
Jean-Louis Tambay, Statistics Canada 100 Tunney's Pasture Driveway Ottawa, Canada, ON K1A 0T6, Jean-Louis.Tambay@statcan.gc.ca

2. METHODOLOGY OF CONFID2

2.1 General description of CONFID2

CONFID2 is a suite of three SAS components, PROC SENSITIVITY, macro SUPPRESS and macro AUDIT, that can create or validate a suppression pattern for tabular economic data (production, employment, financial activity, etc.) at various levels of aggregation. Industrial classification, geographical classification and other classifications could be used as dimensions of the tables.



To run CONFID2 a user must supply four inputs to PROC SENSITIVITY. A microdata file, the hierarchy for each table dimension, the ranges of codes associated with the lowest level of each hierarchy and the sensitivity rule(s) used. The microdata file could be in any format importable into SAS. Ranges define how the microdata records are assembled into tabulation cells. Sensitivity rule(s) are used to calculate cell sensitivity values and identify sensitive cells.

CONFID2 uses the cell suppression technique to protect the confidentiality of respondents' data. This technique consists of primary suppression and residual suppression. PROC SENSITIVITY performs primary suppression and macro SUPPRESS performs residual suppression. The main objective of primary suppression is to identify and suppress confidential cells. A confidential cell is a cell whose value allows a very close estimation of a respondent's contribution. Primary suppression alone does not provide sufficient protection to confidential data since an intruder can calculate either the exact value of a suppressed confidential cell or a very tight range of values for the cell using published internal and marginal cells. That is why after primary suppression CONFID2 performs residual suppression. The main objective of residual suppression is to identify complementary cells (complements) that will also be suppressed in order to protect confidential cells while minimizing the resulting loss of information. If a user modified a suppression pattern produced by CONFID2, then she must run macro AUDIT to verify the validity of the modified suppression pattern. Macro AUDIT also can verify the validity of a suppression pattern produced outside of CONFID2.

2.2 Methodology of PROC SENSITIVITY

PROC SENSITIVITY calculates the totals and sensitivities of tabular cells using microdata. Sensitivity rules are used to calculate cells' sensitivities, and based on these values, to decide which tabular cells are sensitive (i.e. confidential) and to what extent. If a cell's sensitivity is greater than zero, then the cell is sensitive and it requires some protection. Otherwise, the cell is not sensitive. A cell's (positive) sensitivity determines the amount of "protection" that the cell needs by residual suppression. The bigger the cell's sensitivity the greater the protection this sensitive cell requires.

There are 4 sensitivity rules in CONFID2: "N-k", "PQ", Duffett and Arbitrary. All these rules are particular forms of the following general formula:

$$S = \sum_{i=1}^{r} b_i x_i \text{ , } where\ b_1 \geq b_2 \geq \ldots \geq -1 \qquad (1)$$

where S is the cell sensitivity,
b_i are fixed coefficients,
r is the number of contributors to the cell,
x_i are the values of all different contributors to the cell arranged in decreasing order ($x_i \geq 0$).

For efficiency purposes, macro SUPPRESS works with cell-level data rather than respondent microdata. As a result, PROC SENSITIVITY must take extra measures to ensure that certain situations are adequately treated. For example microdata are needed to determine whether two sensitive cells have respondents in common, which will affect their combined sensitivity (in the worst case, making them seem to offer each other "mutual protection" when they do not). PROC SENSITIVITY calculates the sensitivities of unions of sensitive cells, and of unions of sensitive and non-sensitive cells, along one-dimensional lines (rows, columns ...) to identify which of these "aggregates" remain sensitive. Such sensitive aggregates will also be protected by macro SUPPRESS.

PROC SENSITIVITY can process single as well as multiple hierarchical decompositions of any of its dimensions. For example, regions could be sub-divided geographically into Census Divisions, Agricultural Regions and Economic Regions. This will be a triple decomposition of the dimension Region.

As with other SAS procedures, PROC SENSITIVITY can process microdata by groups using “by variables”. For example, a microdata file may contain records for the same enterprises for different years. If we specify variable *year* as a “by variable”, then PROC SENSITIVITY will create and process separate tables for each value of the by variable *year*.

A call for PROC SENSITIVITY looks like any other SAS procedure:

```
PROC SENSITIVITY data=mydata
    hierarchy="industryTotal industry1 industry2; country region1 region 2;"
    range="industry1 industry11 industry12: industry2 industry21 industry22; ; " srule="pq 0.1"
    outcell=main.outcell outconstraint=main.outconstraint;
    id key;
    var income;
    dimension industry region;
run;
```

2.3 Methodology of macro SUPPRESS

The main objective of macro SUPPRESS is to provide the desired protection level to sensitive cells while minimizing the loss of information. To provide the desired protection level to sensitive cells we need to identify complements that will be suppressed during residual suppression. In order to find complements minimizing the loss of information, we need to solve the following LP problem for each sensitive cell X_{sen}.

Matrix formulation of LP problem:

$$\begin{aligned} &\text{Minimize } \mathbf{W'Y + W'Z} \\ &\text{Subject to } \mathbf{CY - CZ = 0} \\ &\qquad \mathbf{0 \le Y, Z \le T/2} \\ &\qquad \mathbf{y_{sen} \ge S_{sen}/2} \\ &\qquad \mathbf{z_{sen} = 0} \end{aligned}$$

where
W: n-dimensional vector of weights (objective function), where n is the total number of cells in a table
Y: n-dimensional vector of positive changes to cell totals
Z: n-dimensional vector of negative changes to cell totals
C: m × n matrix of coefficients, where m = number of linear equations generated by the table
T: n-dimensional vector of cell totals
$\mathbf{y_{sen}}$: the component of the vector **Y** corresponding to cell X_{sen}
$\mathbf{z_{sen}}$: the component of the vector **Z** corresponding to cell X_{sen}
$\mathbf{S_{sen}}$: sensitivity of X_{sen}

The objective function

Among all suppression patterns we would like to select the one that has the smallest number of suppressed cells and for which the total value of all suppressed cells is minimal. Therefore, we would like to minimize an objective function using linear programming (LP) optimization. There are four objective functions in CONFID2 corresponding to different costs of suppression of a cell if it were selected as a complement. The use of different objective functions produces different suppression patterns. The cost of suppressing a sensitive cell is zero since it will be suppressed anyways. The cost of suppressing a cell that had already been selected as a complement to protect another sensitive cell is also zero. The Constant objective function assigns an equal cost of suppression to each non-sensitive cell (for example, for two dimensional problem $w_{ij}=1$). Minimization of this objective function may lead to the suppression of fewer cells with larger totals. The Size objective function assigns a cell’s total value as its cost coefficient ($w_{ij}=t_{ij}$). Its use leads to the suppression of smaller cells. The Information objective function has as coefficients $w_{ij}=[\log(1+t_{ij})]/(1+t_{ij})$. Its minimization leads to the suppression of a small number of large cells, as it is a decreasing function of t_{ij}. The Digit objective function assigns the coefficient $w_{ij}=\log(1+t_{ij})$ and strikes a balance between the Size and Information objective functions. Usually we perform residual suppression twice with different objective functions (this will be explained later).

For the first residual suppression we use the Size, Digit or Constant objective functions, and for the second we use the Information objective function. The following example illustrates the suppression procedure.

Example 1. Let us look at the two dimensional Table 1 of Revenue ($1000s) by Region (first dimension) and Industry (second dimension). NAICS stands for the North American Industrial Classification System. Table 2 is another representation of Table 1.

Table 1

	Total	Industry 1	Industry 2	Industry 3
Total	601	90	300	211
Region 1	140	40	80	20
Region 2	461	50	220	191

Table 2

	Total	Industry 1	Industry 2	Industry 3
Total	$t_{\bullet\bullet}$	$t_{\bullet 1}$	$t_{\bullet 2}$	$t_{\bullet 3}$
Region 1	$t_{1\bullet}$	$t_{11} + 5$	t_{12}	t_{13} - 5
Region 2	$t_{2\bullet}$	t_{21} - 5	t_{22}	$t_{23} + 5$

Suppose that cell X_{23} is a sensitive cell with S_{23}=10. Its *ambiguity* a_{23} ($\equiv S_{23}/2$) is 5, and its upper tolerance is $t_{23} + a_{23} = t_{23} + 5$. To protect this sensitive cell we would like to give it a value at least as large as its upper tolerance. We also could use the lower tolerance ($t_{23} - 5$) because due to the symmetry of the problem we will get the same suppression pattern. Now, we need to balance out the table, say by modifying the following cells: X_{13}, X_{11} and X_{21}. All cells with changed values will be suppressed: they are either sensitive cells or complements. This is how a "safe" suppression pattern is created. Note that the upper bound on cell movements, at one-half of their total value, reflects an assumption that "outsiders" can estimate missing cells to within 50% of their true value. That amount is thus the maximum "protection" that non-sensitive cells can offer.

Let us look at this process using Table 2 from Example 1. Since we do not know in advance which cells will be selected as complements and in which direction their totals will move, let us introduce the notion of a positive change to a total of a cell located in the i^{th} row and j^{th} column (y_{ij}) and a negative change to a total of a cell located in the i^{th} row and j^{th} column (z_{ij}), where z_{ij} is a positive value. Table 2 will generate the following system of equations:

$$\begin{aligned}
&t_{11} + y_{11} - z_{11} + t_{12} + y_{12} - z_{12} + t_{13} + y_{13} - z_{12} = t_{1\bullet} + y_{1\bullet} - z_{1\bullet} \\
&t_{21} + y_{21} - z_{21} + t_{22} + y_{22} - z_{22} + t_{23} + y_{23} - z_{23} = t_{2\bullet} + y_{2\bullet} - z_{2\bullet} \\
&t_{11} + y_{11} - z_{11} + t_{21} + y_{21} - z_{21} = t_{\bullet 1} + y_{\bullet 1} - z_{\bullet 1} \\
&t_{12} + y_{12} - z_{12} + t_{22} + y_{22} - z_{22} = t_{\bullet 2} + y_{\bullet 2} - z_{\bullet 2} \\
&t_{13} + y_{13} - z_{13} + t_{23} + y_{23} - z_{23} = t_{\bullet 3} + y_{\bullet 3} - z_{\bullet 3} \\
&t_{\bullet 1} + y_{\bullet 1} - z_{\bullet 1} + t_{\bullet 2} + y_{\bullet 2} - z_{\bullet 2} + t_{\bullet 3} + y_{\bullet 3} - z_{\bullet 3} = t_{\bullet\bullet} + y_{\bullet\bullet} - z_{\bullet\bullet} \\
&t_{1\bullet} + y_{1\bullet} - z_{1\bullet} + t_{2\bullet} + y_{2\bullet} - z_{2\bullet} = t_{\bullet\bullet} + y_{\bullet\bullet} - z_{\bullet\bullet} \\
&t_{11} + y_{11} - z_{11} + t_{12} + y_{12} - z_{12} + t_{13} + y_{13} - z_{13} + t_{21} + y_{21} - z_{21} + t_{22} + y_{22} - z_{22} + t_{23} + y_{23} - z_{23} = t_{\bullet\bullet} + y_{\bullet\bullet} - z_{\bullet\bullet}
\end{aligned} \qquad (2)$$

Note that last three equations in the system (2) are equivalent; therefore, we will keep only one of the three to avoid redundancies. We can also simplify the system (2) further by bringing everything to the left hand side (LHS) with the opposite sign and cancel some terms, since, for example, $t_{11} + t_{12} + t_{13} = t_{1\bullet}$.

System (3) is a simplified version of system (2).

$$\begin{aligned}
&y_{11} - z_{11} + y_{12} - z_{12} + y_{13} - z_{12} - y_{1\bullet} + z_{1\bullet} = 0 \\
&y_{21} - z_{21} + y_{22} - z_{22} + y_{23} - z_{23} - y_{2\bullet} + z_{2\bullet} = 0 \\
&y_{11} - z_{11} + y_{21} - z_{21} - y_{\bullet 1} + z_{\bullet 1} = 0 \qquad \Leftrightarrow CY - CZ \\
&y_{12} - z_{12} + y_{22} - z_{22} - y_{\bullet 2} + z_{\bullet 2} = 0 \\
&y_{13} - z_{13} + y_{23} - z_{23} - y_{\bullet 3} + z_{\bullet 3} = 0 \\
&y_{1\bullet} - z_{1\bullet} + y_{2\bullet} - z_{2\bullet} - y_{\bullet\bullet} + z_{\bullet\bullet} = 0
\end{aligned} \qquad (3)$$

Since we are creating a suppression pattern to protect cell X_{23} with the variations y_{23} and z_{23}, we will require that z_{23}=0 and y_{23} be at least $a=S_{23}/2$. Should this cell be used as a potential complement for another sensitive cell, the bound of variation would be the same as for other complements: $t_{23}/2$. We will use the Size objective function for this example. Below are the vectors Y, Z, W, T and the matrix C derived from the system of equations (3) for Table 2.

$$Y' = \begin{bmatrix} y_{11} & y_{12} & y_{13} & y_{21} & y_{22} & y_{23} & y_{1\bullet} & y_{2\bullet} & y_{\bullet 1} & y_{\bullet 2} & y_{\bullet 3} & y_{\bullet\bullet} \end{bmatrix}$$

$$Z' = \begin{bmatrix} z_{11} & z_{12} & z_{13} & z_{21} & z_{22} & z_{23} & z_{1\bullet} & z_{2\bullet} & z_{\bullet 1} & z_{\bullet 2} & z_{\bullet 3} & z_{\bullet\bullet} \end{bmatrix}$$

$$W' = \begin{bmatrix} 40 & 80 & 20 & 50 & 220 & 0 & 140 & 461 & 90 & 300 & 211 & 601 \end{bmatrix}$$

$$T' = \begin{bmatrix} 40 & 80 & 20 & 50 & 220 & 191 & 140 & 461 & 90 & 300 & 211 & 601 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

By minimizing **W'Y + W'Z** subject to **CY - CZ = 0; 0 ≤ Y, Z ≤ T/2;** $\mathbf{y_{sen} \geq S_{sen}/2}$ and $\mathbf{z_{sen}=0}$**,** we obtain a solution to the problem in Table 2:

$$Y' = [5 \quad 0 \quad 0 \quad 0 \quad 0 \quad 5 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$$
$$Z' = [0 \quad 0 \quad 5 \quad 5 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$$

In addition to sensitive cell X_{23}, cells X_{11}, X_{13} and X_{21} have changes to their totals, therefore, they are the complements that will be suppressed in order to protect sensitive cell X_{23}.

The sequential approach to finding complements

In the previous section we discussed the creation of a safe suppression pattern for one sensitive cell. Usually, there is more than one sensitive cell to be protected in a table. CONFID2 processes sensitive cells sequentially, one at a time, starting with the sensitive cell that requires the largest amount of ambiguity. In the process of protecting a particular sensitive cell, CONFID2 may partially or fully protect other sensitive cells. The complements for all sensitive cells will be put together to create a suppression pattern for the whole table. Although the procedure is optimal at every step, globally we are not obtaining an optimum solution; therefore, the overall number of small complements is very large and could be reduced by performing a second residual suppression using the Information objective function. This objective function will consider only cells selected during the first residual suppression, since they already satisfy the constraints, and will select larger cells among them as complements to be suppressed, possibly "releasing" some smaller cells to be published.

2.4 Methodology of macro AUDIT

The main objective of macro AUDIT is to verify the validity of a suppression pattern either created by CONFID2 and modified by a user or created outside of CONFID2. To do so for each suppressed cell X_{spr} we will need to calculate a range of possible total values [min t_{spr}, max t_{spr}] by solving two LP problems:

Minimize $\mathbf{v}_{spr}$	Maximize $\mathbf{v}_{spr}$
Subject to **CV = 0**	Subject to **CV = 0**
-LBF*T ≤ V ≤ UBF *T	**- LBF *T ≤ V ≤ UBF *T**

where **V**=Y+Z is an n-dimensional vector of changes to the totals, $\mathbf{v}_{spr}$ is the component of the vector V corresponding to cell X_{spr}, and **LBF** and **UBF** are the factors that reflect the assumption that a potential intruder may estimate a cell total within -LBF*100% and +UBF*100% of the true cell total (e.g. LBF=UBF=½).

When optimal min v_{spr} and max v_{spr} are found, we can calculate a range of possible total values [min t_{spr}, max t_{spr}] using the formulas below:

$$\min t_{spr} = t_{spr} + \min v_{spr} \qquad \max t_{spr} = t_{spr} + \max v_{spr}$$

Note that we are taking into consideration only suppressed cells (sensitive cells and complements); therefore, C, V and T will look as follows:

$$C = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$V' = \begin{bmatrix} v_{11} & v_{13} & v_{21} & v_{23} \end{bmatrix}$$
$$T' = \begin{bmatrix} 40 & 20 & 50 & 191 \end{bmatrix}$$

The solutions to the LP problems are as follows:

$\min V' spr = \begin{bmatrix} 10 & 10 & 10 & 10 \end{bmatrix}$ $\max V' spr = \begin{bmatrix} 10 & 10 & 10 & 10 \end{bmatrix}$

$\min T' spr = \begin{bmatrix} 30 & 10 & 40 & 181 \end{bmatrix}$ $\max T' spr = \begin{bmatrix} 50 & 30 & 60 & 201 \end{bmatrix}$

As we can see, the spread between minimum and maximum values for all suppressed cells is 20, which is greater than the sensitivity S_{23} =10, therefore, CONFID2 created a safe suppression pattern that provides sufficient ambiguity in Table 1 to ensure adequate protection for the sensitive cell X_{23}.

3. COMPARISON OF RESULTS BETWEEN CONFID AND CONFID2

For demonstration purposes we processed four tables from actual surveys at StatCan on a personal computer with the following characteristics: CPU: Intel P4 3.4GHz, RAM-1GB. CONFID2 requires SAS 9.1.3 service pack 4 (or better) as it runs on the powerful SAS/OR® Proc OPTMODEL LP solver. We first looked at the quality of the results produces by the two systems. For this purpose we compared the ratios: (Number of suppressed cells / number of non-zero cells) * 100%. As Table 3 shows, CONFID2 produces slightly lower ratios than CONFID, meaning that it suppresses less. The slight difference may be explained by a better LP process in CONFID2 and by changes in the treatment of sensitive aggregates. In general, the ratios produced by the two systems are very similar, as intended. Second, we compared the execution times for macro SUPPRESS. As seen from Table 3, CONFID2 is at least 25 times faster than CONFID. CONFID2 becomes relatively faster as the number of cells increases. The fifth table, being four-dimensional, cannot be treated by CONFID.

Table 3: Comparison of CONFID and CONFID2

Problem description			Quality results for %Suppress		Execution time for %Suppress	
Name	Dimension	Number of cells	CONFID	CONFID2	CONFID	CONFID2
1	3	664	45.18%	45.18%	6 sec	3 sec
2	3	8,074	23.37%	22.97%	38 min	1.5 min
3	3	9,190	41.64%	41.62%	1 hour 24 min	3 min
4	3	20,688	44.79%	44.12%	13 hours 23 min	13 min
5	4	141,296	n/a	66.79%	n/a	35 hours 22 min

4. CONCLUSION

CONFID2 has been in production at StatCan since May 2009. The preliminary feedback from users is very positive. The system is easy to understand and convenient to work with. Users of CONFID can use a special SAS program to convert the CONFID specification file to CONFID2 specification files.

Some future developments are under consideration. Currently, CONFID2 can create suppression patterns for positive values only; therefore, we would like to conduct research on what is the best methodology for processing negative values. Although CONFID2 is able to produce a suppression pattern for frequency data, it is does not do so efficiently. We would like to improve this. Presently CONFID2 has four predefined objective functions; however, there is a demand for users to be able to specify their own objective functions. CONFID2 has the ability to process very large tables, but it may take a long time. We are investigating ways to reduce the execution time, for example, by using several CPUs or several PCs to run one CONFID2 problem in parallel.

REFERENCES

Jean-Marc Fillion, Ioana Şchiopu-Kratina, Bethany Smith and Jean-Louis Tambay (2007): The SUPPRESS and AUDIT modules of CONFID, Technical Report, Statistics Canada.

Dale Robertson, Ioana Şchiopu-Kratina (1997): "The mathematical basis for Statistics Canada suppression software: CONFID", Survey Methods Section Proceedings from the SSC Annual Meeting, Statistics Canada.