Journal of Official Statistics, Vol. 14, No. 4, 1998, pp. 513–535

Comparison of Systems Implementing Automated Cell Suppression for Economic Statistics

Nancy Kirkendall[1] *and Gordon Sande*[2]

Three automated cell suppression systems were applied to tables from the U.S. Energy Information Administration's Manufacturing Energy Consumption Survey. The four main tables are of moderate size with a complex interrelationship that presents a challenge for confidentiality processing. The article describes, compares, and contrasts the three confidentiality systems, and their application to this complex data set. The result illustrates the importance of auditing proposed suppression patterns, however they are derived, to assure that they offer adequate protection.

Key words: Cell suppression; statistical confidentiality; disclosure control; economic statistics; residual disclosure.

1. Introduction

The most commonly used technique to protect the confidentiality of the data supplied by the respondents in magnitude tabulations of economic censuses and surveys is cell suppression (Schackis 1993, Kirkendall et al. 1994). Cell suppression is the withholding from publication of cells which require protection according to some rule (sensitive cells) along with other cells (complementary cells) to assure that the sensitive cells cannot be derived by manipulating equations specified by the table. Cell suppression has a long history of use and has historically been implemented with manual procedures.

Since the 1970s, major statistical agencies have worked to develop automated procedures as a replacement for the manual procedures. The manual procedures are extremely time consuming for large complex tables and sufficiently error prone that there is no effective guarantee that all sensitive cells are protected. All implementations of cell suppression must balance the competing demands of protecting the confidentiality of the data, providing a usable publication and operating in a timely and efficient manner.

There are two publicly identified automated systems that have been in use for some time. These are the systems from the U.S. Bureau of the Census and from Statistics Canada. There have been repeated suggestions that a comparison of the two systems on

[1] Statistical Policy Office, Office of Management and Budget, New Executive Office Building 10201, 725 17th Street, N.W., Washington, D.C. 20503, U.S.A. (202) 395-7313, Nancy_Kirkendall@omb.eop.gov

[2] Sande and Associates, Inc., 600 Sanderling Court, Secaucus, N.J. 07094, U.S.A. (201) 864-4564, g.sande@worldnet.att.net.

Acknowledgments: This work was carried out while the authors were with EIA. Nancy Kirkendall was a staff member of EIA. Gordon Sande was an Expert/Consultant to EIA. We would like to thank Bob Jewett of USBC for his assistance in operating the USBC system. We would like to thank Mark Schipper of EIA and Susan Bucci of USBC for their subject matter advice. We would like to thank Brian Greenberg, Pat Horning and Laura Zayatz of USBC for their consultation.

a single body of data would be instructive (Cox and Zayatz 1993, Kirkendall et al. 1994). An alternative implementation of the Statistics Canada methods has recently become available. We will report our experiences with these systems in producing tables from live microdata.

Other forms of confidentiality protection are used for other forms of data. Frequency tabulations of demographic data are often protected with random rounding. Microdata releases are often deidentified, sampled and perturbed. The perturbations may take many forms, including top coding and coarsening of classifications as well as error injection and data swapping. Production of economic statistics is only one of many activities of an official statistical agency so cell suppression is only one of many tools used to preserve the confidentiality of their respondents' data.

1.1. Data

The data for this study come from the Manufacturing Energy Consumption Survey (MECS) for 1991 as published in *Manufacturing Consumption of Energy 1991* (EIA 1994). MECS is sponsored by the Energy Information Administration (EIA) of the U.S. Department of Energy, with the survey operations conducted under contract by the U.S. Bureau of the Census (USBC). This study was carried out at the USBC with confidential microdata so there are no examples or detail displayed. All of the tables produced in this study were working tables and could only be reviewed internally at USBC.

MECS has been a triennial survey with 1994 the planned transition year to becoming a biennial survey. Since then budget restrictions have resulted in quadrennial operations only. The data for this study concern components of energy consumption. For the tables used in these examples, the publication is in units of trillion BTU and the microdata is in units of million BTU. (We use quadrillion for 10^{15}, trillion for 10^{12}, billion for 10^9 and million for 10^6.) The data are reported by respondents in physical units and converted to million BTU using standard conversion factors. The conversion to million BTU makes the tables additive, as required for confidentiality analysis. Portions of Tables A1, A3, A4,

Table A1. Total primary consumption of energy for all purposes by census region, industry group and selected industries (estimates in trillion BTU)

Northeast census region

SIC code	Industry groups and industry	Total	Net electricity	Residual fuel oil	Distillate fuel oil	…
…						
25	Furniture and fixtures	7	2	Q	*	…
26	Paper and allied products	W	27	72	*	…
2611	Pulp mills	12	1	2	*	…
2621	Paper mills	228	16	61	W	…
2631	Paperboard mills	W	2	W	Q	…
27	Printing and publishing	23	11	*	1	…
…						

*Estimate less than 0.5. Data are included in higher level totals.

W Withheld to avoid disclosing data for individual establishments. Data are included in higher level totals.

Q Withheld because Relative Standard Error is larger than 50 per cent. Data are included in higher level totals.

Figure 1.1. Portion of Table A1. Selected rows and initial columns only

Table A3. Total primary consumption of combustible energy for nonfuel purposes by census region, industry group, and selected industries (estimates in trillion BTU)

Northeast census region

SIC code	Industry groups and industry	Total	Residual fuel oil	Distillate fuel oil	…
…					
25	Furniture and fixtures	*	0	0	…
26	Paper and allied products	W	0	*	…
2611	Pulp mills	0	0	0	…
2621	Paper mills	*	0	W	…
2631	Paperboard mills	W	0	*	…
27	Printing and publishing	*	0	Q	…
…					

*Estimate less than 0.5. Data are included in higher level totals.
W Withheld to avoid disclosing data for individual establishments. Data are included in higher level totals.
Q Withheld because Relative Standard Error is larger than 50 per cent. Data are included in higher level totals.

Figure 1.2. Portion of Table A3. Selected rows and initial columns only

and A5 from the publication are given in Figures 1.1, 1.2, 1.3, and 1.4. Typical MECS tables are classified by three variables: geography, Standard Industrial Classification (SIC) and fuel type. There are a national total and four geographical regions (Census Regions). For the 1994 and later rounds of MECS, the geographical coding will be more detailed and include a hierarchical disaggregation of the Census Regions. There are an industry total, twenty industry groups (two-digit SIC), and 42 additional major industries (three-digit SIC) or industries (four-digit SIC) specified in 10 of the industry groups. This results in 10 residual industry subtotals which are not presented in the publication. In the portions of tables shown, the residual of SIC 26 is the portion of SIC 26 which is not in SICs 2611, 2621, or 2631. We describe such subtotals as defined but

Table A4. Total inputs of energy for heat, power, and electricity generation by census region, industry group, and selected industries (estimates in trillion BTU)

Northeast census region

SIC code	Industry groups and industry	Total	Net electricity	Residual fuel oil	Distillate fuel oil	…
…						
25	Furniture and fixtures	7	2	Q	*	…
26	Paper and allied products	W	27	72	4	…
2611	Pulp mills	12	1	2	*	…
2621	Paper mills	228	16	61	W	…
2631	Paperboard mills	16	2	W	Q	…
27	Printing and publishing	23	11	*	1	…
…						

*Estimate less than 0.5. Data are included in higher level totals.
W Withheld to avoid disclosing data for individual establishments. Data are included in higher level totals.
Q Withheld because Relative Standard Error is larger than 50 per cent. Data are included in higher level totals.

Figure 1.3. Portion of Table A4. Selected rows and initial columns only

Table A5. Total consumption of offsite-produced energy for heat, power, and electricity generation by census region, industry group, and selected industries (estimates in trillion BTU)

Northeast census region

SIC code	Industry groups and industry	Total	Net electricity	Residual fuel oil	Distillate fuel oil	…
…						
25	Furniture and fixtures	5	2	Q	*	…
26	Paper and allied products	W	32	72	4	…
2611	Pulp mills	5	Q	2	*	…
2621	Paper mills	166	20	61	W	…
2631	Paperboard mills	16	2	W	Q	…
27	Printing and publishing	23	11	*	1	…
…						

*Estimate less than 0.5. Data are included in higher level totals.

W Withheld to avoid disclosing data for individual establishments. Data are included in higher level totals.

Q Withheld because Relative Standard Error is larger than 50 per cent. Data are included in higher level totals.

Figure 1.4. Portion of Table A5. Selected rows and initial columns only

hidden. Their values may often be determined by subtraction. There are a fuel total and eight fuel types.

THE MECS data are further classified into six components by specifying three types of production and two types of consumption. Including these two variables with the original three leads to a five-dimensional data structure. This is more complex data than is typical of economic surveys. A feature of the MECS data which makes confidentiality analysis more complex is that the components are aggregated into four different definitions of

	Component					
	1	2	3	4	5	6
Energy produced						
offsite	x	x				
onsite from nonenergy inputs			x	x		
onsite from energy inputs					x	z
Energy consumed						
as fuel	x		x		x	
for nonfuel purposes		x		x		z
Total Primary consumption for all purposes						
Table A1:	x	x	x	x		
Total Primary consumption for nonfuel purposes						
Table A3:		x		x		z
Total inputs						
Table A4:	x		x		x	
Total consumption of offsite produced energy						
Table A5:	x					

Figure 2. Relationship of components

consumption and published in Tables A1, A3, A4, and A5. The relationship of the six components and the four published MECS tables is shown in Figure 2. This structure can also be given by a set of symbolic equations. In terms of components, the tables are defined by

$$\begin{aligned} A1 &= c1 + c2 + c3 + c4 \\ A3 &= \phantom{c1 +{}} c2 + \phantom{c3 +{}} c4 \\ A4 &= c1 + \phantom{c2 +{}} c3 + \phantom{c4 +{}} c5 \\ A5 &= c1 \end{aligned}$$

and these equations may be *solved* for the components in terms of tables as

$$\begin{aligned} c1 &= \phantom{-A1 + A3 + {}} A5 \\ c3 &= A1 - A3 \phantom{{}+A4} - A5 \\ c5 &= -A1 + A3 + A4 \\ c2 + c4 &= \phantom{-A1 +{}} A3 \end{aligned}$$

From the four published subtotals, it is possible to determine components 1, 3, and 5 and the sum of components 2 and 4. Component 6 is excluded from MECS and defined to be zero. There is a strong relationship between the four published tables. Each of the four tables is a three-way subtotal from a five-way table which is mostly hidden from publication.

A major difference for confidentiality processing between economic data and demographic data is the presence of units with common ownership. These are often called multis, as there are multiple establishments in an enterprise. Multis are typically the large enterprises that both attract the most interest from publication users and require the most care in confidentiality protection by the publication producers. The degree of common ownership across either SIC or geographical disaggregations in typical economic publications is moderate. MECS is unusual as there is a higher degree of common ownership across the fuel type disaggregations. Within an SIC and geographical location, energy intensive establishments tend to have similar fuel use patterns so there will be similar ownership patterns for the fuel type disaggregations.

1.2. Automated systems

An automated system is both a theoretical specification of what to do and an operational specification of how to do it. In practice the use that a system is put to may not match the restrictions that result from its theoretical foundations. The systems that we consider are in operational configurations for real publications in production environments. There are many features in common as the systems are solving the same problem.

1.2.1. Tabulation

Tabulating microdata to produce values of cells in a publication is a standard activity. It is often done by a process called rolling up in which totals are determined by adding up their components. Determining the value of the sensitivity function of cells is more difficult. The sensitivity function uses both the value of the cell and the value of the largest respondents contributing to the cell. The identity of the largest respondents may change under aggregation when there are multis present. There are examples of national firms which are smaller than each of their regional competitors in each region but are still the largest firm nationally.

1.2.2. Suppression

The suppression stage takes the information about the sensitive cells and the sensitive aggregates from the tabulation stage and calculates a pattern of complementary suppressions. The computation is an application of mathematical programming, or optimization, methods. All the systems use a similar sequential heuristic in which the most sensitive cell is processed first in a single cell protection step, and then the next most sensitive cell and so on until all cells have been processed. This is the greedy heuristic that is widely used in mathematical programming.

The single cell protection step acts by temporarily changing the value of the cell to be protected to the boundary between too close and not too close an approximation. The resulting unbalanced table is then rebalanced to provide a self consistent table completion and all of the modified cells are then withheld as complementary suppressions. The rebalancing is restricted so that the modified values are consistent with the external user knowledge about the cells. There are many patterns of rebalancing possible. The possible patterns are given a figure of merit by specifying numerical coefficients and mathematical programming used with the figure of merit as an objective function to find the preferred configuration. Differing specifications of the objective function will lead to differing styles of choice of complementary suppressions.

1.2.3. Processing

The practical suppression activity will require some amount and various kinds of support. The users may want to specify which cells are hidden, are required to be published or are available for withholding. They may also want to modify the suppression pattern presented by the automated systems. The problem may be partitioned into subproblems for various operational reasons. Superfluous suppressions may be removed by use of alternate operating modes.

1.2.4. Auditing

Auditing finds the lower and upper bounds on the values that a withheld cell may assume and still be consistent with the table. The range of values may also be of use to end users of the table. The model of end user knowledge will affect the interpretation of the calculated bounds on the withheld cells. If the information available is only the final suppression pattern, so the model of user knowledge will only be that the withheld values are positive, the audit can only verify the absence of inadvertent exact residual disclosures. Such a weak model of user knowledge will be an underestimate of the knowledge of an interested industry analyst.

If the information available includes both the cell values and the cell sensitivities from the tabulation stage, so that the model of user knowledge can be approximated for all withheld cells, the audit can verify the absence of inadvertent approximate residual disclosures. Such a strong model of user knowledge provides an independent verification of the suppressions actions. This would only be possible within the secure environment of a statistical agency.

Auditing often requires more computing time than suppression. Suppression has larger systems but can be partitioned. It need only examine sensitive cells and may bypass sensitive cells that are protected *in passing*. Auditing must examine all withheld cells and does not benefit from partitioning.

2. USBC System

The USBC system may be described as a network theory based system. There are various articles on the application of network theory to cell suppression (Carvalho et al. 1994, Cox 1995). Only the earliest developments of the theory have been implemented in full production systems. A full production system must address various issues outside the restrictions of network theory. These issues may include the possibility of hierarchy in both classification variables such as SIC by geography, non-hierarchical classification variables such as geography with metropolitan areas, higher dimensions or sensitive aggregates due to low respondent counts or units with common ownership.

The USBC system is in its third version. The first version, named INTRA, is identified with Cox (1980). It selected at least two cells for suppression in every row, or column, that had suppressions with the intent of reducing the total value of suppressed cells. The selection criteria did not account for the simultaneous effects of rows and columns and thus required an audit procedure to identify defective patterns that resulted. The second version, unnamed, is identified with Hemmig (Jewett 1993). It used a network optimization code within a sequential heuristic to choose complementary suppressions with the intent of reducing the total value of suppressed cells. The analysis was for a single two-way table set up as a transportation problem. The method is self auditing as it cannot lead to defective patterns in two dimensions. A batched analog of backtracking was used for the between table analysis. Both versions used a cell by cell definition of sensitivity.

The third and current version, unnamed, was developed by Jewett (1993). It uses the same network optimization code within a sequential heuristic for two-way tables related by a single hierarchical variable set up as a transshipment problem. It uses the same batched analog of backtracking for the between table analysis not included in the single hierarchical variable. A capacity reduction technique is used to extend the definition of sensitivity to pairs of cells. A third classification variable can be processed as a set of unrelated one-way problems and the results require an audit procedure to identify defective patterns.

2.1. *Tabulation*

The USBC system does not have a tabulation component. It requires that the microdata be tabulated by a tabulation program and specialized results provided for further processing. Within USBC, such tabulation programs are available for each survey and a new one would be either a new development or a modification of an existing one. The output of the tabulation program is presumed to be self-consistent and any error checking and diagnosis would depend upon the tabulation program. For each cell the tabulation program provides the total value, the number of respondents and the values and identifiers for the two largest respondents (Jewett 1993). The sensitivity function is evaluated in the suppression stage from this information. The limited information provided by the tabulation stage restricts the available sensitivity rules to either the *n-k%* rules for *n* of 1 or 2, the *p%* rules or rules which can be constructed by combining these two forms. This limitation would be a restriction for some users. If the microdata is weighted, the weights are used to determine the cell value. The difference between the unweighted value of a respondent and the weighted value of that respondent acts as an imputed respondent which cannot be one of the large respondents.

The result of the tabulation is a fixed field flat file with one record for each cell. One of the fixed fields is a processing status code which can be used to specify various actions, including prepublication, and presuppression. Modification of the processing status code can also be done directly using a text editor.

For the MECS data, the USBC tabulation program divided all cell values by 1,000 to keep the grand total within the range of integer values. The grand total of energy consumption is about 20 quadrillion BTU or 20 billion microdata units of million BTU. A computer integer in a computer word of 32 bits can have only about 10 digits so the change in units was required. Some small cells would be rounded to zero as their values are less than 500, except that this special case results in a rounded value of 1. Zero valued cells can not be complements as the absence of respondents is assumed to be known to all data users.

2.2. *Suppression*

The USBC system evaluates the sensitivities of the cells as part of the suppression stage. The sensitivity rule in use must be specified. The USBC does not release the parameters of its rule.

The USBC system deals with tables which do not add up exactly by introducing a correction term. When the source of not adding is independent rounding, the correction term is small and this is an effective way of avoiding a potentially awkward technical problem. When the source of not adding is a missing classification, such as the SIC 26 residual in the MECS example, this can cause incorrect processing as the correction term does not have its largest respondents identified and is treated as being not sensitive. In practice the missing classification may be sensitive and require complementary suppression to protect it.

The confidentiality protection model given in various USBC technical reports is a cell by cell model (Zayatz 1992). If a row of a table had two sensitive cells, they would serve as complements for each other unless there was great disparity in their sizes. If each of these cells had a single respondent, the combined total would be known and would be a one or two respondent miscellaneous aggregation. For the case of one repeated respondent, the total would be known by all table users. For the case of two respondents, either respondent would then be able to determine the value of the other. The current version of the USBC system does not follow the model given in the earlier USBC technical reports, but rather adjusts the apparent size, called the protection capacity, of a cell to deal with this situation. The protection capacity is calculated by examining the pooling of the cell being protected and the cell being evaluated for its protection capacity. The result is that the protection capacity available from a cell is different at each step of the sequential heuristic.

The capacity reduction technique addresses the two-cell miscellaneous aggregate problem but leaves other problems unsolved. For a single repeated respondent, the sensitivity of the two-cell aggregate is greater than that of either cell and could be assigned to a single cell, to avoid under-protection, but that may cause problems for the column of that single cell. A problem also arises when there are several sensitive cells in a row. The pooling of all the cells may or may not be sensitive but each proposed complement may have no protection capacity when examined. If the proposed complements have respondents in common, the analysis would be correct and the pool would be sensitive. If the proposed complements do not have respondents in common, the analysis would be conservative and the pool might not be sensitive. The capacity reduction technique does not fully deal with the several-cell miscellaneous aggregate problem as it must implicitly make the conservative estimate.

The sensitivity of the aggregates may be underestimated because of the limited information on respondents. The information available makes the evaluation of the sensitivity of the pooling of two cells incorrect in some cases. A configuration in which cell x has respondent A of value 100 and cell y has respondents A, B, and C of values 20, 40, and 40, respectively, leads to an incorrect pooling of $x + y$ with A of (incorrect) value 100, B of value 40 and the remaining 60 being assigned to smaller respondents as the identity of A is lost from y. The correct pooling yields a value of 120 for A, 40 for B and 40 for smaller respondents. For a *2–75%* rule, for example cell x is sensitive as one respondent is 100%, cell y is sensitive as two respondents are 80% but $x + y$ is not sensitive as two respondents seem to be 70%. Under the correct pooling, $x + y$ is sensitive as two respondents are 80%.

2.3. *Processing*

Problem partitioning is a central part of the design of the USBC system. All problems are partitioned into smaller components which are processed as network problems. The smaller components are then recombined by the backtracking process. The USBC system processes the four three-way tables as a single task. Each table has 3,285 cells, processed as five geographic layers of 657 cells. The layer is a single disaggregation for fuel type by a hierarchical disaggregation for SIC which does not require any backtracking. The geographic dimension is treated as multiple one-way disaggregations that may cause backtracking. The layers are not large when compared to other problems. The processing time was below fifteen minutes on the VAX cluster at USBC.

The output is a revised set of processing status codes in a file with a format of the input file. The revised processing status codes are used by other programs in the USBC computing environment. Another of the output files is a listing of all the complementary cells which can be used to explain why any cell has been suppressed. There is an auxiliary program to display this file. Experienced users may search the file directly using a text editor.

2.4. *Auditing*

There is a postprocessing stage to evaluate the sensitivity of the aggregations formed of all the suppressed cells in a row, or column or slice, of a table. These aggregations should not be sensitive. There is an audit capability which is rarely used as it is too time-consuming in operation (Kirkendall et al. 1994). The run times suggested were considerably higher than those observed with CONFID and ACS. Discussion with the USBC staff revealed that the USBC system audit used a convenience interface to a standard professional linear programming package. The convenience interface had the effect of requiring the linear programming package to repeat both the Phase I and Phase II calculations for every lower or upper bound obtained. For a moderate-sized problem, the Phase I calculation in which the linear program package finds its first basic feasible solution may take several hundred pivotal exchanges with an equivalent amount of work in Phase II to find the optimal solution. To proceed from an optimal solution for one objective function to an optimal solution of a modified objective function typically involves a small number of pivotal exchanges. In this case the use of the convenience interface of the linear program package had a very high cost for repeated solutions.

2.5. Additional comments

The system is operated by its developers. New surveys are dealt with by changing the parameter files and incorporating changes into the existing system. Alternate objective functions are available by modification of the cost coefficients of the system. The system is operational on DEC VAX computers at USBC and can only be moved to other systems with some effort as it uses capabilities of the VAX/VMS Record Management System that would require a database management system on other systems. An implementation away from the USBC would require a suitable tabulation program for support as well.

The USBC system is available without cost from its developers. Since the completion of this study a version of the system suitable for operation on a PC has become available.

3. CONFID and ACS Systems

CONFID and ACS may be described as general simplex theory based systems. The basic technology of the two systems is the same. One may view CONFID as the prototype for ACS. The differences are those one would expect to find between the first and second version of the same system. The main differences are in software engineering and usability issues and the treatment of dimensionality.

The Statistics Canada system, named the Confidentiality Studies Software but commonly known as CONFID, was developed by Sande (1984). It was developed as a research prototype and is in use at Statistics Canada (Robertson 1993). Separate versions of CONFID can process either two non-hierarchical classification variable problems or three hierarchical classification variable problems. The analysis to choose complementary suppressions uses a general simplex optimization code within a sequential heuristic. The definition of sensitivity is extended to multiple cells by the use of miscellaneous aggregations. There is a choice of objective functions to approximate either least count or least total value of suppressed cells as well as the Burg (1967) entropy, which has become the default compromise between the other two objective functions. There is another objective function used to release suppressions as part of a two-stage cleanup process. Other components of CONFIG include tabulation, auditing of suppression patterns, adjustment of tables with suppressions to be additive and multiple single purpose utilities.

The newly available system, named Automated Cell Suppression (ACS), was developed by Sande (1995). It builds on experience with CONFID but is independent of Statistics Canada. ACS can process problems with from one to seven non-hierarchical classification variables. The basic techniques are the same as those of CONFID but there are considerable revisions to the grouping of functions. The two stages of the cleanup process and the multiple steps of a segmented problem can be executed as a single computer task. Other functions of ACS include tabulation, auditing of suppression patterns, adjustment of tables with suppressions to be additive, self-consistent completion of tables, various utility actions and extraction of values from spreadsheets.

3.1. Tabulation

CONFID and ACS provide tabulation components. The microdata is expected to be a fixed field flat file. For CONFID the field positions are fixed and the field contents are numerical values for the classification variables, the respondent identifier and the data value. For ACS the field positions may be specified and the classification variables may include alpha-numerical values. This would allow mnemonic codes to be used for geographical region or fuel type in the MECS data. If there are coding errors in the data, CONFID will diagnose the presence of aggregations which do not add correctly. The user must then find the coding errors by other means. ACS will diagnose coding errors in the records as unknown code values.

The microdata for both tabulation programs is assumed unweighted. If the microdata is weighted, then as part of the data preparation two records for each respondent would be constructed. One record would have the value for the respondent and the other record would have the value for the imputed respondent with an identifer indicating that this is an anonymous respondent.

There are also miscellaneous aggregates that are sensitive. The miscellaneous aggregates address the issue of the presence of the same economic respondents in several cells. In CONFID, a miscellaneous aggregate will be defined by aggregating the sensitive cells in a row, column or slice if certain conditions are met. The conditions are that the miscellaneous aggregate must be sensitive and the estimate of its sensitivity obtained from bounds applied to all components should include both positive and negative values. The bounds will always include positive values. For two components, if one component is very sensitive and the other so small that it could not provide protection, the bounds will have only positive values. The determination of the sensitivity of the aggregate directly from the microdata allows the influence of repeated respondents to be accounted for. In ACS, the analysis is extended to detect the case where non-sensitive cells in the same row, column or slice etc., can be included in the pooled cells and the result will still be sensitive.

The sensitivity rule in use must be specified to the tabulation programs. There are options to provide an *n-k%* rule, a *p%* rule or a general linear weighting of the largest respondents. These have a minor restriction as only the five largest respondents are identified. The tabulation output is the value of all the defined cells and a numerical value of the sensitivity of the sensitive cells, transformed to be an upper tolerance according to the definitions used by these systems. The output files contain similar information in formats intended for use as input to other components of the respective systems.

3.2. Suppression

CONFID and ACS use the sensitivity information supplied by their tabulation components. The tabulation components are specialized for their tasks and provide information that requires no further manipulation. The method is self-auditing as the use of the simplex method in the internal step cannot lead to a defective pattern in any number of dimensions. Both systems use double precision internally to represent large values exactly. All aggregates are required to be the exact sum of their disaggregates as part of their error checking. They could have dealt with the MECS microdata even if it had not been in units of BTU rather than million BTU.

3.3. Processing

Utility programs are available to indicate prepublication, presuppression and other suppression options. CONFID has multiple single purpose utilities. ACS has a single multiple purpose utility.

The three-dimension limit for CONFID meant that it could only process the tables as separate three-dimensional tables, and could not process them as interrelated tables. The ACS processing was done with the five-way table. The microdata available provided access to components 1, 3, and 5 and the sum of components 2 and 4. If the components were classified with two hierarchical variables for production and consumption, there would be 15 classifications or subtotals present. There would be nine distinct non-zero classifications or subtotals and six zero or redundant classifications or subtotals for the four published subtotals. If the components were classified with one non-hierarchical variable for component number there would be seven distinct non-zero classifications or subtotals and no zero or redundant classifications or subtotals for the four published subtotals. The three hidden tables would correspond to components 3 and 5 and to the overall total. The one non-hierarchical classification variable form was used.

CONFID and ACS can process a single 3,285 cell table in under 15 minutes on MECS's mini computer. The default ACS run has an initial suppression pass and a second cleanup pass. The corresponding sequence of actions in CONFID takes several computer tasks to do the two passes and to apply the utilities several times. The times are similar except for the additional input and output needed to process several tasks. The utility programs can convert the output to a fixed field flat file with status codes which was used for doing cross-comparisons with the USBC system. The same flat file would be used with the USBC publication and dissemination systems.

The mini computer available for processing the MECS data has been in service for some time. It would be considered slow for other purposes but is adequate for its one and only client. It appears to be between five and ten times slower than the VAX cluster in use with its many clients.

The five-way table has 22,995 cells with 11,655 hidden cells and 11,340 cells in the publication. This problem is sufficiently large that it was segmented to reduce the execution time on MECS's mini computer. The initial segment selected only the total and the 20 SIC industry groups and all other variables for 6,615 cells. Two passes were used to identify and then clean up the suppressed and releasable cells and they were flagged as such. The next segment selected the initial segment and about 10 additional SICs from several industry groups. This bigger problem, with the suppression and release results of the initial segment applied as presuppression and prerelease conditions, was solved and the results flagged for the newly selected cells. Another three segments each selected the initial segment and about 10 different additional SICs from several different industry groups. These bigger problems, with the results of the initial segment applied, were solved and the results flagged for the newly selected cells. With the initial segment and four additional segments which were differing partial disaggregations of the initial segment, all cells had been processed. The result was five steps, each of two passes, specified by five selection commands in a single computer task. The corresponding sequence of actions in CONFID is ten steps each of a suppression pass and several utility tasks. The execution time for MECS's mini computer was under 150 minutes.

CONFID and ACS report their suppression actions as they proceed, so that one can explain the use of complements. When a sensitive cell is protected *in passing* while protecting another sensitive cell, the complements may be more generous than technically required. ACS has a separate pass available to report the complements needed for each sensitive cell in the final pattern. The listing indicates both the complements of each sensitive cell and the sensitive cells of each complement. In this separate pass, no sensitive cell is processed *in passing* so the execution time may exceed the time for the initial determination of the complements.

3.4. Additional comments

Statistics Canada has made CONFID available under restricted conditions to some U.S. statistical agencies (EIA and USBC) upon request (SAIC 1985). The restrictions include no technical support. CONFID is operated by its end users. New surveys require new control statements for a general purpose system. CONFID has operated on a variety of computers including IBM mainframes, workstations such as DEC VAX with VMS, Sun with Unix or HP with Unix, or microcomputers such as IBM/PC compatible or Macintosh with MPW.

ACS is available under license from Sande and Associates. The copy used for these studies was made available to EIA for demonstration use. ACS is operated by its end users. New surveys would require preparation of new control commands for a general purpose program. ACS has operated on a variety of computers including workstations such as DEC VAX with VMS, Sun with Unix or HP with Unix, or microcomputers such as IBM/PC compatible or Macintosh with MPW.

Since the completion of this study ACS has extended its audit capacity by finding isolated cells and determining their values with simple methods. This would calculate many of the hidden MECS cells. Additional suppression heuristics intended to reduce some forms of over-suppression have been added. In sparse tables a small sensitive cell protected at a late processing stage may require a large complementary suppression that would have been beneficial in the protection of other cells at earlier processing stages. Look ahead and reordering heuristics for the repeated single cell protection methods may lower the over-suppression in such cases. For small tables a suppression method that protects all sensitive cells simultaneously, rather than by repeatedly protecting single cells, is also available.

4. Comparison Study

The comparison study followed the common style of doing some simple tasks first. An initial simple task was to audit the released publication from its machine readable versions. This provided an opportunity to learn the structure of the publication. All of the tables were dealt with as they were presented in the publication. The auditing of three-dimensional tables was straightforward, with only the technical difficulty of the independent rounding of the table entries. The next step was to reproduce the tables from the microdata. The organization of the microdata matched the released tables. As this microdata had previously been tabulated for publication, the transformations necessary to carry questionnaire responses to tabulation microdata were already defined and implemented.

All three systems were used to calculate suppression patterns in many three-dimensional tables. Cross-comparisons to ensure that consistent interpretations of the data were being used were done and some minor problems sorted out. The SIC residual subtotals which are hidden in the publication were not initially specified for tabulation in the USBC system. This was noticed in comparisons of the outputs for the systems and a revised tabulation specification was prepared and this problem was corrected.

Whenever a data collection and a processing system are brought together for the first time it is common to discover that some aspect of some specification is not fully met. The symptoms of this may vary from diagnostic error messages to unexpected or erroneous results. The MECS microdata has a simple coding structure which caused no trouble. The number of negative data values for net electricity and other (usually net steam) was slightly higher than expected. Absolute values of the data were used in this study.

The subject matter advisors had indicated that the major interest in the publication was in the tables which reported the four definitions of consumption. It became obvious that the *Net Electricity* column of Tables A1 and A4 were identical as they used the same microdata field as tabulation input. The *Net Electricity* column in Table A3 is absent, or structurally zero. Explanations from the subject matter advisors and their suggested reading of the publication's preface material brought the structure discussed above to our attention. This was a forceful reminder of the great importance of understanding the structure of the data. The simple initial tasks and the learning phase of the comparison study gave way to the more complex task of dealing with the four strongly interrelated tables. This would turn into the most interesting part of the comparison and is the bulk of what we report here. In the common way with many studies, the *first 90 per cent* of the study took the *first half* of the time and this *final 10 per cent* took the *final half* of the time.

The USBC system represented this additional structure by using inequalities between tables which were related by being subsets. The duplicated column in Tables A1 and A4 was represented by placing the same column in two aggregation specifications. CONFID and ACS imposed stricter conditions which would not permit the direct placement of the same column in two aggregation specifications. To achieve the same effect required the use of the higher-dimensional structure in which this was a logical consequence of the classification structure and the structural zeroes. CONFID could not represent the higher-dimensional structure and could not be used for the joint processing of the interrelated tables. ACS could represent the higher dimensional structure directly. This illustrates the differing operational styles of the systems. The USBC system is tolerant of incomplete or inconsistent user specifications as the specifications are presumed to be correct. CONFID and ACS insist upon logically complete and consistent user specifications and perform checks to ensure that this is true.

The comparison was done in two stages. The initial stage used only the three-dimensional structure of the data. It was more intensive with cross-checks between the three systems. The cleanup modes of CONFID and ACS were used to remove possible superfluous suppressions. Table A1 was used as the primary example. The final stage used the five-dimensional structure of the data jointly in the four tables. It used ACS to check for exact and approximate residual disclosures for ACS and the USBC system.

4.1. *Using the USBC system*

The post-processing step of the USBC system showed that ACS had found two complements for two sensitive cells in a row but that the aggregation of the four cells was still sensitive. The default analysis of the sensitive cells plus one potential complement was inadequate for this data. The analysis was extended to consider four potential complements by changing an option value to specify searching four potential complements. The conservative assumptions implicit in the capacity reduction used by the USBC system are appropriate for the high degree of common ownership seen for the fuel type disaggregation of the MECS data. The USBC system audit was not applied to the results of either the USBC system or ACS.

4.2. *Using ACS*

The ACS audit of the ACS suppression patterns found no problems. During the initial stage, an ACS audit of a USBC system suppression pattern detected an inadvertent approximate residual disclosure of a miscellaneous aggregation. Using that USBC system suppression pattern as a starting point for an ACS suppression resulted in some additional complements to fully protect the miscellaneous aggregation and the release of 76 of 876 complements in the ACS cleanup pass.

During the final stage, for the five-way tables the ACS audit runs for the ACS-produced patterns produced the lower and upper bounds on all suppressed values in the five-way table, including all the hidden values used as complements. The result of the run showed no audit exceptions for the ACS produced suppression pattern.

During the final stage, for the USBC system produced patterns all hidden values were assumed to be suppressed. There were about 605 audit exceptions for the USBC system produced suppression pattern. About 485 of the audit exceptions were exact or approximate residual disclosures of miscellaneous aggregations. About 45 audit exceptions were approximate residual disclosures of sensitive cells in the publication. About 75 of the audit exceptions were exact residual disclosures of cells in the publication. Of these, 15 were of sensitive cells and 60 were of non-sensitive cells intended to be complements. It was verified that the USBC system had operated as intended and that the disclosures were the result of using the joint structure of the four tables and not just some previously unnoticed fault. In an earlier five-way table trial run there had been more cells suppressed but fewer audit exceptions, so the conclusion was that the between table protection was being provided by the fortunate alignment of suppressions into favorable configurations. When the five-way table structure was being considered, more cells suppressed meant a better chance of having a fortunate configuration of cells.

5. Comparisons

Data collectors and providers consider it their duty to publish as much information as possible subject to protecting the confidentiality of individual respondents' data. The two commonly suggested ways to accomplish this are either to minimize the number of suppressed cells using the philosophy that cells are equally important or to minimize the total value of suppressed cells using the philosophy that the larger cells are of greater interest to users. In fact, most data providers would like to accomplish both of these objectives simultaneously.

In suppression software these alternatives are implemented by using an objective function that provides a weight for each suppressed cell. To minimize the number of suppressed cells one makes use of a *constant* objective function, weighting all cells equally. To minimize the total value of suppressed cells, one makes use of a *size* objective function, weighting each cell by its value. To minimize the information in the suppressed cells one uses a *digits* objective function, weighting each cell by the logarithm of its value. CONFID and ACS provide options allowing users to select different objective functions. The USBC system currently uses only the *size* objective function.

Figures 3.1, 3.2, and 3.3 illustrate the effect of the objective function on the number of suppressed cells and the total value of suppressed cells using ACS (Kirkendall et al. 1996). The cells have been classified as being marginal or internal cells. When viewed jointly, Tables A1 and A4 have no internal cells as Tables A3 and A5 are internal to Table A1 and Table A5 is internal to Table A4. There is much multiple counting in the total value comparison. The internal cells do not have this multiple counting. The total value suppressed for internal cells can be compared to the value of the grand totals given in Figure 4. The grand totals are for the absolute values of the microdata and will not agree exactly with the publication as some data items, such as net electricity, can be negative. The published tables also reflect some additional values obtained from other EIA sources for refineries (SIC 29).

These figures illustrate that the objective functions produce the expected results. Using the *constant* objective function, Figure 3.1, one tends to have fewer suppressed cells, but a somewhat larger suppressed value. Using the *size* objective function, Figure 3.3, the number of suppressed cells tends to be larger, but the total value suppressed is smaller. The *digits* objective function tries to control both value and count. As a result, the statistics for Figure 3.2 tend to fall between the other two.

Another interesting comparison is the amount of suppression due to internal cells, versus marginal totals. With the *constant* objective function approximately 25 per cent of the suppressed cells were internal, but they averaged about 23 per cent of the total table value. For the *size* objective function approximately 30 per cent of the suppressed cells were internal, but they constituted 40 per cent of the total table value. The *digits* objective function again falls between the other two, with 29 per cent of the suppressed cells internal,

		Count and value of suppressed cells		
		All cells	Margin cells	Internal cells
Jointly	Count	1,317	1,190	127
	Value	226,274	221,960	4,313
Table A1	Count	497	364	133
	Value	103,852	96,794	7,057
Table A3	Count	102	90	12
	Value	20,958	19,845	1,113
Table A4	Count	355	263	92
	Value	67,711	63,498	4,213
Table A5	Count	363	248	115
	Value	33,750	30,551	3,199

Figure 3.1. Constant objective function for ACS

		Count and value of suppressed cells		
		All cells	Margin cells	Internal cells
Jointly	Count	1,373	1,198	175
	Value	171,449	165,435	6,014
Table A1	Count	479	335	144
	Value	71,494	66,718	4,776
Table A3	Count	121	100	21
	Value	23,918	21,686	2,231
Table A4	Count	319	210	109
	Value	39,002	35,750	3,252
Table A5	Count	454	300	154
	Value	37,034	33,251	3,782

Figure 3.2. Digits objective function for ACS

accounting for about 37 per cent of the total table value. It appears that as the weight for a cell increases from one to its cell value in the objective function, the cell suppression algorithm is more likely to pick internal table cells. Marginal cells are larger than the internal cells which they contain.

The USBC system results with the *size* objective function are provided in Figure 3.4 (Kirkendall et al. 1996). The USBC system suppresses fewer cells than the ACS system with the *size* objective function (see Figure 3.3) at the expense of total value suppressed. The system appears to favor internal cells: about 38 per cent of the suppressed cells are internal, and these contribute to an average of about 49 per cent of the table total value. The results presented here have not been corrected for the 75 exact residual disclosures revealed in the final stage auditing or for the additional suppressions that would be required to correct these problems.

A subjective evaluation of the varying objective functions produced a confirmation of the strong preference for preserving the cells associated with energy intensive industries

		Count and value of suppressed cells		
		All cells	Margin cells	Internal cells
Jointly	Count	1,812	1,590	222
	Value	105,016	99,134	5,882
Table A1	Count	609	385	224
	Value	30,821	24,893	5,927
Table A3	Count	159	138	21
	Value	19,046	16,528	2,517
Table A4	Count	464	292	172
	Value	24,214	20,363	3,851
Table A5	Count	580	379	201
	Value	30,934	27,570	3,364

Figure 3.3. Size objective function for ACS

		Count and value of suppressed cells		
		All cells	Margin cells	Internal cells
Jointly	Count	2,664	2,267	397
	Value	90,073	82,534	7,538
Table A1	Count	759	382	377
	Value	16,739	10,839	5,900
Table A3	Count	163	132	31
	Value	13,008	10,370	2,638
Table A4	Count	913	539	374
	Value	40,561	33,974	6,586
Table A5	Count	829	463	366
	Value	19,763	14,863	4,900

Figure 3.4. USBC system

and heavily used fuel types. Compared to many economic surveys, MECS is a relatively specialized survey and has been designed to deal with the energy intensive industries. It is natural that this strong subject matter content should be reflected in the design of the suppression patterns as well. (The original intent of this study had been to compare the results of the USBC system with those of ACS and to compare the results of both systems under several iterations of subject matter advice. Both ACS and USBC systems provide

Cell values	
Jointly	18,804
Table A1	17,662
Table A3	3,503
Table A4	15,301
Table A5	10,765

Figure 4. Grand total

mechanisms to incorporate user preferences into the process of selecting a usable suppression pattern. The extent of the study was reduced due to lack of extended access to the subject matter advisors resulting from the disruptions to recover from the partial government closures in the period from October 1995 to February 1996.)

6. Further Work

6.1. Integration of rounding and protection

The comparisons did not consider the impact of rounding the publication to trillion BTU. In the publication there are symbols of $*$ to represent small values which would round to zero but which are not exactly zero. A $*$ could be as much as 500,000 million BTU. It is a practical observation of business statistics that small values and sensitive values often go together.

A small sensitive cell might have a value of 260,000 million and we would consider it protected if all interval estimates were less precise than the interval 220,000 to 300,000 million. If we suppress before rounding, we would report a W although a $*$ would represent an interval sufficiently wide to protect the cell. When there is a W, we would expect to find complementary Ws even if the $*$ would be adequate protection. It would be possible to have a $*$ as a non-sensitive marginal cell with the sensitive components as Ws. These would also be known to be $*$s. This illustrates that the rounding introduces an interval of uncertainty for all the cells, and this uncertainty may be larger than the uncertainty that we are trying to introduce to protect the confidentiality of the data. However, these two sources of uncertainty are not coordinated. The small complements of small sensitive cells in this analysis illustrates this lack of coordination of the two sources of uncertainty.

Rounding leads naturally to ranges. It is commonly understood that a value rounded to be 3 may be in the range of 2.5 to 3.5. When 4 has been rounded to be even, we understand that it represents a range of 3 to 5. We do not have a simple convention for representing a range of 2.5 to 4.5, although the symbols in use for ratings by consumer magazines provide useful examples. It is often not clear whether 10 means the range of 9.5 to 10.5 or of 5 to 15, although we might use 10 and 1$*$ for the two possible rounding bases of 1 and 10. This shows the problem of indicating the base when we are rounding to alternative bases.

A subtotal value of 1$*$, (5 to 15) with two components of values 2 (1.5 to 2.5) and 8 (7.5 to 8.5) would be recognized as being a range of 9 to 11 upon closer analysis. A subtotal of value 1$*$ with ten components of value 1 would not permit such an refinement. These are two examples of interval arithmetic. All this is a reminder of the close relationships between rounding, ranges, and confidentiality. We should also remember that ranges can be from sources of error such as sampling or accounting for differing fiscal years of establishments.

6.2. Protection by ranges

We could use error ranges to provide confidentiality protection. If the error ranges act to refine the confidentiality protection ranges, we could introduce complementary protection ranges in the same way that we introduce complementary cell suppressions. We notice that complementary protection ranges correspond to increasing the rounding base. Some experiments have suggested that the need for complementary protection ranges is pleasantly small when error ranges are available. High level aggregates can be given with no ranges for the many existing uses of such aggregates without impairing the usefulness of ranges for lower-level cells. These experiments have not addressed the problem of how to report the ranges or even how to pick the initial protection ranges. We would not want the protection ranges to be centered so that their midpoints were the values we were hoping to protect. The ranges of required protection for cell suppression have traditionally been centered on the true value, but are only displayed as part of internal confidential working tables or as non-confidential examples. The ranges from rounding give no hint of what the original internal value may have been, but are often wider than the required protection ranges. This returns us to the problem of representing a range such as 2.5 to 4.5 and reporting the rounding base that applies to it.

Many users might be initially confused by the presence of ranges. The request to *just give me the number* comes readily to mind and may be partially addressed by giving popular high level aggregates without ranges. For lower level cells the availability of self-consistent completions of the tables would reduce this problem as well as the existing problem of suppressed cells. It is easy to forget that many users view a suppressed cell as a complete absence of any useful information when it really represents considerable information readily available technically. The user education problem for ranges would be similar to that for suppressed cells, which appears to have been rarely attempted or achieved. Some of the user education may also be useful for data producers. By the time users understand lower and upper bounds on cells and self-consistent completions of tables for cell suppression, they will view ranges as a minor variation on the same techniques.

7. Review and Summary

The USBC system was developed to replace manual operations by equivalent computer operations. There had been little flexibility in the manual operations and little was provided in the computer operations. Low computer cost was an early and dominant consideration in the development process. The initial development was restricted in the analytical resources available and various limitations are a result of the design decisions which limited the information available from the tabulations. The various versions were directed at resolving operational difficulties.

CONFID was developed as a research prototype intended to further the understanding of the problems of publishing tables while protecting the confidentiality of the data and to identify the requirements of a production system. The initial requirements were for flexibility in attempting to solve various problems without the need for operational integration in solving those various problems. The research prototype has been in operation for many years. ACS uses the same methods as CONFID but packages them very differently. Some standard operating sequences observed with CONFID, such as the suppression stage followed by a cleanup stage or the multiple stages of a segmented problem, have been greatly simplified or made automatic defaults.

MECS is a complex problem of only a moderate size. The USBC system can process the three-way tables, but gives no guarantees. It uses inequalities to represent the relationship between the main tables and fails to adequately protect the data. ACS processed all of the three- and five-way tables and provided audits to verify the results.

We see that small problems can be handled quickly by the systems which were being compared here. Small size for automation can be moderate size for manual operations. For the small problems, the computer time is more influenced by issues such as amount of data read, output produced and other system issues than by the asymptotic computer science computational complexity of the core algorithms. Issues of development cost, flexibility and usability become more important when there are many small problems. For big problems, the computational complexity of the core algorithms can become an issue although we see that big problems are also likely to be more complex and the correctness and completeness of the core algorithms can become an issue as well.

Problem partitioning which leads to lower execution costs is achieved by backtracking for the USBC system and by segmentation for CONFID and ACS. The USBC method requires the backtracking techniques to preserve the network structure of problems which also leads to many small problems and lowers the computational cost. Miscellaneous aggregations do not appear to be possible within the network structure. There are various USBC technical reports which document the over-suppression that can result from the backtracking technique (Sullivan 1993). However, in this example the USBC system performed quite well in terms of the number of cells suppressed.

The grouping of hierarchically related tables into bigger problems is intended to improve the choice of complements at the expense of execution cost. The CONFID and ACS methods may be either carried out in a single step of low cost for small problems or as several segmented steps for lowering of what would otherwise be a big cost for a high problem. Backtracking is required by the USBC system for all but simple problems. Segmentation is optional for CONFID and ACS. Realistic timing comparisons would be based on problem partitioning for all systems. For complex technologies, it is easy to make major performance compromises, as illustrated by the high cost of the use of convenience interfaces for the USBC system audit.

8. Conclusions

It is important to understand the structure of the data. The dimensionality of the data may not be initially apparent and may not match the apparent dimensionality of its standard presentations.

All the systems operate more quickly than analysts can review their output so the variations in execution cost are not a qualitative comparison attribute. The ability to vary possible suppression patterns to reflect subject matter is an important benefit beyond the speed and accuracy of the automated suppression patterns. Much as subject matter concerns can influence the sampling design of the survey, subject matter concerns should also influence the suppression design of the publication.

The value of the ability to independently verify the successful operation of a suppression program was illustrated a number of times in this project. Of particular note were the identification of the disclosures in the USBC suppression pattern by the ACS audit and the USBC post-processing of an early ACS run identified the need to change the specification of an input parameter. The former demonstrate the utility of an independent audit capability. The latter demonstrates the utility of a simple audit to verify that the aggregates of all suppressed cells in a row, column or layer are not sensitive according to the sensitivity rule.

9. References

Burg, J. (1967). Maximum Entropy Spectral Analysis. Paper presented at the 37th Meeting of the Society of Exploration Geophysicists, Oklahoma City.

de Carvalho, F.D., Dellaert, N.P., and de Sanches Osorio, M. (1994). Statistical Disclosure in Two-dimensional Tables: General Tables. Journal of the American Statistical Association, 89, 1547–1557.

Cox, L. (1980). Suppression Methodology and Statistical Disclosure Control. Journal of the American Statistical Association, 75, 377–385.

Cox, L. (1995) Network Models for Complementary Cell Suppression. Journal of the American Statistical Association, 90, 1453–1462.

Cox, L. and Zayatz, L. (1993). Setting an Agenda for Research in the Federal Statistical System: Needs for Statistical Disclosure Limitation Procedures. Proceedings of the Section on Government Statistics, American Statistical Association, 121–126.

EIA (1994). Manufacturing Consumption of Energy 1991. Publication DOE/EIA-0512(91), Washington, DC: Energy Information Administration.

Jewett, R. (1993). Disclosure Analysis for the 1992 Economic Census. Unpublished manuscript, Washington, DC: Economic Programming Division, U.S. Bureau of the Census.

Kirkendall, N. et al. (1994). Report on Statistical Disclosure Limitation Methodology. Statistical Policy Working Paper 22, Washington, DC: Office of Management and Budget.

Kirkendall, N. et al. (1996). Report on EIA-Census Evaluation of Disclosure Limitation Methods. Unpublished manuscript, Washington, DC: Office of Statistical Standards, Energy Information Administration.

Robertson, D. (1993). Cell Suppression at Statistics Canada. Proceedings of the 1993 Annual Research Conference, U.S. Bureau of the Census, 107–131.

SAIC (1985). Preserving Confidentiality in Energy Publications: Introduction to Using CONFID. Unpublished manuscript, Washington, DC: Office of Statistical Standards, Energy Information Administration.

Sande, G. (1984). Automated Cell Suppression to Preserve Confidentiality of Business Statistics. Statistical Journal of the United Nations ECE, 2, 33–41.

Sande, G. (1995). ACSSuprs. Unpublished manuscript, Secaucus, NJ: Sande and Associates.

Schackis, D. (1993). Manual on Disclosure Control Methods. Manuscript, Luxembourg: Eurostat.

Sullivan, C. (1993). A Comparison of Cell Suppression Methods. Unpublished manuscript ESMD-9301, Washington, DC: Economic Statistics Methods Division, U.S. Bureau of the Census.

Zayatz, L. (1992). Linear Programming Methodology for Disclosure Avoidance Purposes at the Census Bureau. Proceedings of the Section on Survey Research Methods, American Statistical Association, 679–684.

Received August 1997
Revised September 1998