How to Prevent Cell Perturbation Procedures from Becoming Data Falsification Procedures (Incorporating Quality Measures in Tabular Data Protected by Cell Perturbation Methods) A Practical Solution

Ramesh A Dandekar

U. S. Department of Energy, Washington DC 20585
Ramesh.dandekar@eia.doe.gov

Abstract. Statistical agencies routinely publish analysis reports by using survey data summarized in tabular format. The analysis reports often provide industry specific regional and sub-regional fluctuations in the reported tabular values by comparing them with the data collected in previous survey cycles. Based on the information presented in the tabular data, data users often make their own inferences about the changes in the market conditions for their own immediate geographical areas. To continue to meet these requirements, newly proposed cell perturbation-based tabular data protection methods such as controlled tabular adjustment (Dandekar 2001), and a micro data level noise addition method need to have a strategy in place to ensure that the data users do not confuse cell perturbation error with fluctuations in published cell values attributed to other sources. To achieve this objective, we propose a simple mechanism to continue to safeguard data quality and to provide the most accurate information to data users in the tabular format protected by cell perturbation methods. *Our proposed method with minor modifications could be used to convey the relative standard errors (RSE) associated with tabular format estimates derived from sampled survey data.*

1 Introduction

Statistical agencies routinely publish analysis reports by using survey data summarized in tabular format. The analysis reports often provide industry specific regional and sub-regional fluctuations in the reported tabular values by comparing them with the data collected in previous survey cycles. The comparative evaluation helps analysts to determine long and short term trends at local, regional and national levels. Similarly, such a comparison allows analysts to observe unusual fluctuations in reported tabular values over time resulting from either the policy changes or due to changes in the market conditions. Statements such as, "*Coal consumption decreased in the electric power sector by 1.0 percent, the coking coal sector by 2.0 percent, and the other industrial sector by 1.4 percent.*" or "*Data show that total generation in the electric power sector (electric utilities and independent power producers) in the United States increased slightly in 2006.*" are routinely used in statistical publications.

Based on the information presented in the tabular data, data users often make their own inferences about the changes in the market conditions for their own immediate geographical areas. To meet such a stringent requirement, newly proposed cell perturbation-based tabular data protection methods such as controlled tabular adjustment (Dandekar 2001), and a micro data level noise addition method need to have a unified strategy in place to ensure that the data users do not confuse cell perturbation error introduced by these methods to protect sensitive tabular cells with fluctuations in published cell values resulting from a variety of other sources, including sampling and non-sampling errors . A practical mechanism is required to be in place to continue to safeguard data quality and to provide the most accurate information to data users in the tabular format using these methods.

2 A Practical Solution

To convey the accuracy and the quality of ***table cell specific*** information to the external data users, we propose using a table cell imbedded quality indicator for all tabular data protected by the cell perturbation methods. Our proposal advocates:

- To use a very small percentage change in the perturbed cell value (such as 0.01% or even smaller) as a threshold for quality acceptance criteria.
- To publicly disclose the threshold quality acceptance value for perturbed cells in a footnote to all the public use tables protected by cell perturbation methods.
- To suppress the selected number of right most digits from the perturbed tabular cells with perturbation error in excess of the threshold acceptance value. This is to inform data users the extent of error associated with the published perturbed cell value.

We propose the following equation to determine the number of right most digits to suppress:

Right most digits to suppress = Integer Value[$\log_{10}$ (2.0 * Abs {change}) + 1.0]

The multiplier of two in the equation above is introduced to eliminate the possibility of indirect disclosure of sensitive cells by using external pattern auditing procedures.

It is important to note that attempts by the external data users to estimate the values for the suppressed digits will never result in the disclosure (exact or statistical) of sensitive cell values. At best, the estimate for the suppressed digits will result in external data users recovering the perturbed value for the sensitive cell. As a result, withholding the number of digits for only one cell in a given row or a column should not be considered as a potential disclosure problem arising from this procedure.

3 An Illustrative Example

Let us assume that the change made to a cell value within a plus or minus 0.5 percent of the true cell value is considered to be as good as the true cell value and therefore is published in its entirety.

Example 1:

Cell Value: 172 Change: -8

Percent Cell Value Change: -8 / 172 * 100 = -4.65%

Number of cell value digits to suppress = Integer Value [$\log_{10}$ (2.0 * 8) + 1.0] = 2

Adjusted cell value = 172 - 8 = 164.

Published cell value = 1xx.

Example 2:

Cell Value: 3840 Change: -8

Percent Cell Value Change: -8 / 3840 * 100 = -0.21 % …. *Within threshold acceptance level*

Published cell value = Adjusted cell value = 3840 – 8 = 3832

4 Example Using a Real Life Table Structure

To demonstrate the effectiveness of our proposed "quality preserving" solution, we use the same real life table structure of moderate hierarchical and linked complexity used in the technical paper, "Comparative Evaluation of Four Different Sensitive Tabular Data Protection Methods Using a Real Life Table Structure of Complex Hierarchies and Links", (Dandekar 2007). The table consists of eight two-dimensional cross sections linked in the four-dimensional space. Appendix A shows a layout for one of the eight two-dimensional cross sections.

We use the table layout from Appendix A to display the proposed published format output from our procedure. The rows in each of the eight cross sections provide geographical details consisting of hierarchical structure. The table is populated with the non-real synthetic micro data using the same procedure described in that paper (Dandekar 2007). The p percent rule with p=10% was used to identify sensitive cells. To demonstrate the effectiveness of the micro data level noise addition method as compared to the CTA method, the micro data used as an input to create the table was perturbed by using a bimodal normal distribution with mean value of 0.9 and 1.1, and standard deviation of 0.005. ***Both methods (CTA and noise) provided same level of protection to sensitive tabular cells. However as explained in Dandekar 2007 paper, the noise method makes excessive adjustments to non-sensitive tabular cells when compared to the CTA method.***

To illustrate our approach, we use *0.01% cell perturbation error* as a threshold for quality acceptance criteria. **Table 1** shows a part of the table from Appendix A belonging to PAD District I. The table is protected by using a LP-based controlled tabular adjustment (CTA) procedure targeted towards larger tabular cells (Dandekar 2001). The first four columns in the table show the published format table cell values after our proposed right most digit suppression logic is applied to the cell values that are adjusted by using the CTA procedure. The last four columns in the table show the required cell specific adjustments from the CTA procedure. As a part of our proposal, the footnote to the table clearly identifies the threshold quality acceptance criteria used in the table. The withheld digits in the cell values are shown by symbol 'x'.

Table 1

CTA Based Table 01 [TARGET LARGE CELLS]   REGULAR

	DTW	RACK	BULK	TOTAL	< ---- Adjustments			---->
United States	188xxx*	218xxx*	169xxx*	577xxx*	-130.A	113.A	-61.A	-78.A
PAD District I	646xx*	73xxx*	65xxx*	203xxx*	-8.A	69.A	-143.A	-82.A
Subdistrict IA	253xx*	287xx*	16952	71046	-8.A	8.A	0.	0.
Connecticut	6258	1494	1700	9452	0.	0.	0.	0.
Maine	3936	4719	4429	13084	0.	0.	0.	0.
Massachusetts	1xx*	38xx*	0	4012	-8.w	8.A	0.	0.
New Hampshire	7879	0	3188	11067	0.	0.	0.	0.
Rhode Island	1748	6224	3976	11948	0.	0.	0.	0.
Vermont	5321	12503	3659	21483	0.	0.	0.	0.
Subdistrict IB	19417	165xx*	22xxx*	582xx*	0.	48.A	-61.A	-13.A
Delaware	6978	24xx*	4272	137xx*	0.	48.A	0.	48.A
District of Columbia	2253	5070	11338	18661	0.	0.	0.	0.
Maryland	3311	1836	1xxx*	6xxx*	0.	0.	-60.w	-60.A
New Jersey	6875	0	144	7019	0.	0.	0.	0.
New York	0	648	7xx*	13xx*	0.	0.	-39.w	-39.A
Pennsylvania	0	6539	47xx*	113xx*	0.	0.	38.A	38.A
Subdistrict IC	19894	277xx*	26xxx*	73xxx*	0.	13.A	-82.A	-69.A
Florida	0	10857	18xx*	126xx*	0.	0.	-17.A	-17.A
Georgia	9961	0	0	9961	0.	0.	0.	0.
North Carolina	2268	72xx*	8xxx*	17xxx*	0.	13.A	-65.A	-52.A
South Carolina	1195	5887	7582	14664	0.	0.	0.	0.
Virginia	3560	0	3625	7185	0.	0.	0.	0.
West Virginia	2910	3751	4815	11476	0.	0.	0.	0.

*Lower Digits Withheld. Table Cell Perturbation Error Exceeds 0.01%

Based on the table content, the user knows that the PAD district total is within a range from 203,001 to 203,999 with an average value of 203,500. The uncertainty range of the PAD district total value can be estimated to be plus or minus 500 / 203,500 * 100 = 0.245%. Large numbers such as 203,xxx are used in practice in the thousands of measurement units and therefore are useful to most users with partial digits suppressed without uncertainty range computations. Similar computations could be performed for other tabular cells with withheld numbers of digits on an as-needed basis. When the table cell values are available in their entirety, users know for sure that the perturbation error is less than 0.01%. To allow comparative evaluation of two data perturbation methods, Appendix B shows the published table values based on our proposed method when the micro data level noise addition method is used as a statistical disclosure control strategy instead of CTA method.

5 Comparative Assessment – Targeting Smaller vs Larger Cells

The current research related to the CTA method is mostly based on using larger non-sensitive cells to counter balance the adjustments required to protect sensitive tabular cells. However, the practitioners of the CTA method have a wide variety of options available through a selection of an appropriate objective function to select table cells for adjustments. As an option, the CTA procedure could be targeted toward smaller non-sensitive cells to counter balance the adjustments made to protect sensitive table cells. **Table 2** (a part of the table corresponding to PAD District I from Appendix C) shows the outcome from the CTA procedure when a cost function proportional to the cell value is used in the LP-based procedure. By using smaller cells for adjustments allows the larger cells to be published in their entirety, while for the small value cells the number of digits are withheld after these cells are adjusted to protect sensitive cells. This could be a preferred option for many applications.

Table 2

CTA Based Table 01 [TARGET SMALL CELLS] REGULAR

	DTW	RACK	BULK	TOTAL	< ---- Adjustments			---->
United States	188668	218471	170021	577160	0.	0.	0.	0.
PAD District I	64625	72994	65620	203239	0.	0.	0.	0.
Subdistrict IA	25314	28780	16952	71046	0.	0.	0.	0.
Connecticut	62xx*	15xx*	1700	9452	-8.A	8.A	0.	0.
Maine	3936	4719	4429	13084	0.	0.	0.	0.
Massachusetts	1xx*	38xx*	0	4012	8.w	-8.A	0.	0.
New Hampshire	7879	0	3188	11067	0.	0.	0.	0.
Rhode Island	1748	6224	3976	11948	0.	0.	0.	0.
Vermont	5321	12503	3659	21483	0.	0.	0.	0.
Subdistrict IB	19417	16493	22335	58245	0.	0.	0.	0.
Delaware	6978	2xxx*	4xxx*	13650	0.	85.A	-85.A	0.
District of Columbia	2253	5070	11338	18661	0.	0.	0.	0.
Maryland	3311	1xxx*	1xxx*	6226	0.	-60.A	60.w	0.
New Jersey	6875	0	1xx*	70xx*	0.	0.	-14.A	-14.A
New York	0	6xx*	8xx*	14xx*	0.	-25.A	39.w	14.A
Pennsylvania	0	6539	4718	11257	0.	0.	0.	0.
Subdistrict IC	19894	27721	26333	73948	0.	0.	0.	0.
Florida	0	10857	1847	12704	0.	0.	0.	0.
Georgia	9961	0	0	9961	0.	0.	0.	0.
North Carolina	2268	72xx*	84xx*	17958	0.	39.A	-39.A	0.
South Carolina	1195	5887	7582	14664	0.	0.	0.	0.
Virginia	3560	0	3625	7185	0.	0.	0.	0.
West Virginia	2910	37xx*	48xx*	11476	0.	-39.A	39.A	0.

*Lower Digits Withheld. Table Cell Perturbation Error Exceeds 0.01%

6 Comparative Assessment – Threshold Quality Acceptance Value

A variety of threshold values could be used depending on the quality requirements from the data users. Here we demonstrate the outcome when the threshold quality acceptance criterion is lowered to 1.0%. In **Table 3** (a part of the table corresponding to PAD District I from Appendix D), we target larger cells for the adjustments. In **Table 4** (a part of the table corresponding to PAD District I from Appendix F), we target smaller cells for the adjustments. As expected, fewer cells are flagged by using the lower quality acceptance criteria. Appendix E shows the published table values based on our proposed method when the micro data level noise addition method is used as a statistical disclosure control strategy.

Table 3

CTA Based Table 01 [TARGET LARGE CELLS]   REGULAR

	DTW	RACK	BULK	TOTAL	< ---- Adjustments			---->
United States	188538	218584	169960	577082	-130.A	113.A	-61.A	-78.A
PAD District I	64617	73063	65477	203157	-8.A	69.A	-143.A	-82.A
Subdistrict IA	25306	28788	16952	71046	-8.A	8.A	0.	0.
Connecticut	6258	1494	1700	9452	0.	0.	0.	0.
Maine	3936	4719	4429	13084	0.	0.	0.	0.
Massachusetts	1xx*	3848	0	4012	-8.w	8.A	0.	0.
New Hampshire	7879	0	3188	11067	0.	0.	0.	0.
Rhode Island	1748	6224	3976	11948	0.	0.	0.	0.
Vermont	5321	12503	3659	21483	0.	0.	0.	0.
Subdistrict IB	19417	16541	22274	58232	0.	48.A	-61.A	-13.A
Delaware	6978	24xx*	4272	13698	0.	48.A	0.	48.A
District of Columbia	2253	5070	11338	18661	0.	0.	0.	0.
Maryland	3311	1836	1xxx*	6166	0.	0.	-60.w	-60.A
New Jersey	6875	0	144	7019	0.	0.	0.	0.
New York	0	648	7xx*	13xx*	0.	0.	-39.w	-39.A
Pennsylvania	0	6539	4756	11295	0.	0.	38.A	38.A
Subdistrict IC	19894	27734	26251	73879	0.	13.A	-82.A	-69.A
Florida	0	10857	1830	12687	0.	0.	-17.A	-17.A
Georgia	9961	0	0	9961	0.	0.	0.	0.
North Carolina	2268	7239	8399	17906	0.	13.A	-65.A	-52.A
South Carolina	1195	5887	7582	14664	0.	0.	0.	0.
Virginia	3560	0	3625	7185	0.	0.	0.	0.
West Virginia	2910	3751	4815	11476	0.	0.	0.	0.

*Lower Digits Withheld. Table Cell Perturbation Error Exceeds 1.0%

Table 4

CTA Based Table 01 [TARGET SMALL CELLS]   REGULAR

	DTW	RACK	BULK	TOTAL	< ---- Adjustments			---->
United States	188668	218471	170021	577160	0.	0.	0.	0.
PAD District I	64625	72994	65620	203239	0.	0.	0.	0.
Subdistrict IA	25314	28780	16952	71046	0.	0.	0.	0.
Connecticut	6250	1502	1700	9452	-8.A	8.A	0.	0.
Maine	3936	4719	4429	13084	0.	0.	0.	0.
Massachusetts	1xx*	3832	0	4012	8.w	-8.A	0.	0.
New Hampshire	7879	0	3188	11067	0.	0.	0.	0.
Rhode Island	1748	6224	3976	11948	0.	0.	0.	0.
Vermont	5321	12503	3659	21483	0.	0.	0.	0.
Subdistrict IB	19417	16493	22335	58245	0.	0.	0.	0.
Delaware	6978	2xxx*	4xxx*	13650	0.	85.A	-85.A	0.
District of Columbia	2253	5070	11338	18661	0.	0.	0.	0.
Maryland	3311	1xxx*	1xxx*	6226	0.	-60.A	60.w	0.
New Jersey	6875	0	1xx*	7005	0.	0.	-14.A	-14.A
New York	0	6xx*	8xx*	1446	0.	-25.A	39.w	14.A
Pennsylvania	0	6539	4718	11257	0.	0.	0.	0.
Subdistrict IC	19894	27721	26333	73948	0.	0.	0.	0.
Florida	0	10857	1847	12704	0.	0.	0.	0.
Georgia	9961	0	0	9961	0.	0.	0.	0.
North Carolina	2268	7265	8425	17958	0.	39.A	-39.A	0.
South Carolina	1195	5887	7582	14664	0.	0.	0.	0.
Virginia	3560	0	3625	7185	0.	0.	0.	0.
West Virginia	2910	37xx*	4854	11476	0.	-39.A	39.A	0.

*Lower Digits Withheld. Table Cell Perturbation Error Exceeds 1.0%

Similarly, in **Table 5** and **Table 6** we use the most stringent quality acceptance criteria possible of 0.0 percent to demonstrate the outcome from the CTA by using the larger cells for adjustments (**Table 5**) and by using the smaller cells for adjustment (**Table 6**). *This quality acceptance criterion could be the desirable criterion to maintain the integrity of tabular data protected by cell perturbation methods.*

Table 5

CTA Based Table 01 [TARGET LARGE CELLS] REGULAR

	DTW	RACK	BULK	TOTAL	< ---- Adjustments			---->
United States	188xxx*	218xxx*	169xxx*	577xxx*	-130.A	113.A	-61.A	-78.A
PAD District I	646xx*	73xxx*	65xxx*	203xxx*	-8.A	69.A	-143.A	-82.A
Subdistrict IA	253xx*	287xx*	16952	71046	-8.A	8.A	0.	0.
Connecticut	6258	1494	1700	9452	0.	0.	0.	0.
Maine	3936	4719	4429	13084	0.	0.	0.	0.
Massachusetts	1xx*	38xx*	0	4012	-8.w	8.A	0.	0.
New Hampshire	7879	0	3188	11067	0.	0.	0.	0.
Rhode Island	1748	6224	3976	11948	0.	0.	0.	0.
Vermont	5321	12503	3659	21483	0.	0.	0.	0.
Subdistrict IB	19417	165xx*	22xxx*	582xx*	0.	48.A	-61.A	-13.A
Delaware	6978	24xx*	4272	137xx*	0.	48.A	0.	48.A
District of Columbia	2253	5070	11338	18661	0.	0.	0.	0.
Maryland	3311	1836	1xxx*	6xxx*	0.	0.	-60.w	-60.A
New Jersey	6875	0	144	7019	0.	0.	0.	0.
New York	0	648	7xx*	13xx*	0.	0.	-39.w	-39.A
Pennsylvania	0	6539	47xx*	113xx*	0.	0.	38.A	38.A
Subdistrict IC	19894	277xx*	26xxx*	73xxx*	0.	13.A	-82.A	-69.A
Florida	0	10857	18xx*	126xx*	0.	0.	-17.A	-17.A
Georgia	9961	0	0	9961	0.	0.	0.	0.
North Carolina	2268	72xx*	8xxx*	17xxx*	0.	13.A	-65.A	-52.A
South Carolina	1195	5887	7582	14664	0.	0.	0.	0.
Virginia	3560	0	3625	7185	0.	0.	0.	0.
West Virginia	2910	3751	4815	11476	0.	0.	0.	0.

*Lower Digits Withheld. Table Cell Perturbation Error Exceeds 0.0%

Table 6

CTA Based Table 01 [TARGET SMALL CELLS] REGULAR

	DTW	RACK	BULK	TOTAL	< ---- Adjustments			---->
United States	188668	218471	170021	577160	0.	0.	0.	0.
PAD District I	64625	72994	65620	203239	0.	0.	0.	0.
Subdistrict IA	25314	28780	16952	71046	0.	0.	0.	0.
Connecticut	62xx*	15xx*	1700	9452	-8.A	8.A	0.	0.
Maine	3936	4719	4429	13084	0.	0.	0.	0.
Massachusetts	1xx*	38xx*	0	4012	8.w	-8.A	0.	0.
New Hampshire	7879	0	3188	11067	0.	0.	0.	0.
Rhode Island	1748	6224	3976	11948	0.	0.	0.	0.
Vermont	5321	12503	3659	21483	0.	0.	0.	0.
Subdistrict IB	19417	16493	22335	58245	0.	0.	0.	0.
Delaware	6978	2xxx*	4xxx*	13650	0.	85.A	-85.A	0.
District of Columbia	2253	5070	11338	18661	0.	0.	0.	0.
Maryland	3311	1xxx*	1xxx*	6226	0.	-60.A	60.w	0.
New Jersey	6875	0	1xx*	70xx*	0.	0.	-14.A	-14.A
New York	0	6xx*	8xx*	14xx*	0.	-25.A	39.w	14.A
Pennsylvania	0	6539	4718	11257	0.	0.	0.	0.
Subdistrict IC	19894	27721	26333	73948	0.	0.	0.	0.
Florida	0	10857	1847	12704	0.	0.	0.	0.
Georgia	9961	0	0	9961	0.	0.	0.	0.
North Carolina	2268	72xx*	84xx*	17958	0.	39.A	-39.A	0.
South Carolina	1195	5887	7582	14664	0.	0.	0.	0.
Virginia	3560	0	3625	7185	0.	0.	0.	0.
West Virginia	2910	37xx*	48xx*	11476	0.	-39.A	39.A	0.

*Lower Digits Withheld. Table Cell Perturbation Error Exceeds 0.0%

7 Basic Properties of Proposed Practical Solution

The proposed technical solution allows data users to estimate the accuracy and the uncertainty range associated with all tabular cells protected by the perturbation methods. The uncertainty range estimated by the external data users is comparable (on the wider side) to the uncertainty range introduced by the cell perturbation method.

Withholding the number of digits for only one sensitive cell in a given row or a column does not cause the statistical disclosure of the perturbed cell. This is apparent from Table 3. In that table, for the State of Massachusetts the number of digits are withheld for only one cell. The cell is sensitive with a true cell value of 172. However, for the external users the uncertainty range is estimated to be 125.52 to 202.48 [from (4012 – 1.01 * 3848) to (4012 – 0.99 * 3848)] with an expected value of 164.

Unlike conventional cell suppression methods, estimates are available for all tabular cells. By using the appropriate cost function among many available cost functions, the CTA outcome could be targeted to maximize the overall quality of the tabular data to meet the requirements of a majority of data users.

The method could be considered as a hybrid of various tabular data protection methods such as conventional cell suppression methods, partial cell suppression, selective rounding, and the interval estimation method.

8 Future Work

In this paper we have established a basic framework to convey table cell specific information on the quality and the accuracy of tabular data protected by various perturbation methods. The concept of "threshold quality acceptance value" introduced in this paper is flexible enough to accommodate a wide variety of table structures requiring different quality expectations. The paper, however, does not address combined effects of intentional perturbation errors on other sources of survey errors attributed to sampling and non-sampling activities. To achieve that objective, a "joint threshold quality acceptance value" will required to be developed to combine various components of conventional survey errors with intentional perturbation error.

The "rightmost digit suppression method" from this paper makes indiscriminant use of the multiplier two to provide conservative uncertainty estimates to external data users. The multiplier is introduced to eliminate the possibility of indirect disclosure of sensitive cells by using external pattern auditing procedures. In an attempt to release more digits of information for the perturbed cells, a selective use of the multiplier of two on only sensitive table cells and for small value cells will need to be evaluated carefully to prevent indirect statistical disclosure by automated mathematical procedures.

The complete 72 pages of output from the test example used in this paper is available for all eight two-dimensional cross sections, three scenarios (CTA targeted at large cells, noise method and CTA targeted at small cells) and three threshold values from the URL http://mysite.verizon.net/vze7w8vk/t4344_All_per.pdf

References

Dandekar R. A. (2001) "Synthetic Tabular Data: A Better Alternative To Complementary Data Suppression - Original Manuscript Dated December 2001". Energy Information Administration, U. S. Department of Energy. Also available from CENEX-SDC Project International Conference, PSD2006, Rome, Italy, December 13-15, 2006, Companion CD Proceedings ISBN: 84-690-2100-1.

Dandekar R. A. and Cox L. H. (2002), Synthetic Tabular Data: An Alternative to Complementary Cell Suppression, 2002. Manuscript, Energy Information Administration, U. S. Department of Energy.

Dandekar, R.A (2003), Cost Effective Implementation of Synthetic Tabulation (a.k.a. Controlled Tabular Adjustments) in Legacy and New Statistical Data Publication Systems, working paper 40, UNECE Work session on statistical data confidentiality (Luxembourg, 7-9 April 2003)

Dandekar Ramesh A. (2004), Maximum Utility-Minimum Information Loss Table Server Design for Statistical Disclosure Control of Tabular Data, pp 121-135, Lecture Notes in Computer Science, Publisher: Springer-Verlag Heidelberg, ISSN: 0302-9743, Volume 3050 / 2004**,** Title: Privacy in Statistical Databases: CASC Project International Workshop, PSD 2004, Barcelona, Spain, June 9-11, 2004.

Dandekar Ramesh A. (2005), "Complementary Cell Suppression Software Tools for Statistical Disclosure Control - Reality Check" , Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality (Geneva, Switzerland, 9-11 November, 2005)

Dandekar Ramesh A. (2007), "Comparative Evaluation of Four Different Sensitive Tabular Data Protection Methods Using a Real Life Table Structure of Complex Hierarchies and Links", working paper 17, UNECE Work session on statistical data confidentiality (Manchester, United Kingdom, Dec 17-19, 2007)

Evans T., Zayatz L., Slanta J. (1998),"Using Noise for Disclosure Limitation of Establishment Data", Journal of Official Statistics, Vol.14, No.4, 1998. pp. 537-551

Appendix A

CTA Targeted Towards Larger Tabular Cells – 0.01% Perturbation Error Acceptable

CTA Based Table 01 [TARGET LARGE CELLS]   REGULAR

	DTW	RACK	BULK	TOTAL	< ---- Adjustments ---->			
United States	188xxx*	218xxx*	169xxx*	577xxx*	-130.A	113.A	-61.A	-78.A
PAD District I	646xx*	73xxx*	65xxx*	203xxx*	-8.A	69.A	-143.A	-82.A
Subdistrict IA	253xx*	287xx*	16952	71046	-8.A	8.A	0.	0.
Connecticut	6258	1494	1700	9452	0.	0.	0.	0.
Maine	3936	4719	4429	13084	0.	0.	0.	0.
Massachusetts	1xx*	38xx*	0	4012	-8.w	8.A	0.	0.
New Hampshire	7879	0	3188	11067	0.	0.	0.	0.
Rhode Island	1748	6224	3976	11948	0.	0.	0.	0.
Vermont	5321	12503	3659	21483	0.	0.	0.	0.
Subdistrict IB	19417	165xx*	22xxx*	582xx*	0.	48.A	-61.A	-13.A
Delaware	6978	24xx*	4272	137xx*	0.	48.A	0.	48.A
District of Columbia	2253	5070	11338	18661	0.	0.	0.	0.
Maryland	3311	1836	1xxx*	6xxx*	0.	0.	-60.w	-60.A
New Jersey	6875	0	144	7019	0.	0.	0.	0.
New York	0	648	7xx*	13xx*	0.	0.	-39.w	-39.A
Pennsylvania	0	6539	47xx*	113xx*	0.	0.	38.A	38.A
Subdistrict IC	19894	277xx*	26xxx*	73xxx*	0.	13.A	-82.A	-69.A
Florida	0	10857	18xx*	126xx*	0.	0.	-17.A	-17.A
Georgia	9961	0	0	9961	0.	0.	0.	0.
North Carolina	2268	72xx*	8xxx*	17xxx*	0.	13.A	-65.A	-52.A
South Carolina	1195	5887	7582	14664	0.	0.	0.	0.
Virginia	3560	0	3625	7185	0.	0.	0.	0.
West Virginia	2910	3751	4815	11476	0.	0.	0.	0.
PAD District II	76xxx*	62147	54xxx*	193xxx*	-71.A	0.	126.A	55.A
Illinois	4128	0	0	4128	0.	0.	0.	0.
Indiana	45xx*	0	38xx*	8459	-14.A	0.	14.A	0.
Iowa	1149	4196	4216	9561	0.	0.	0.	0.
Kansas	11xxx*	10330	2xxx*	24xxx*	-57.A	0.	112.A	55.A
Kentucky	5826	2787	6523	15136	0.	0.	0.	0.
Michigan	2022	0	6668	8690	0.	0.	0.	0.
Minnesota	6400	3694	1332	11426	0.	0.	0.	0.
Missouri	5915	10385	3934	20234	0.	0.	0.	0.
Nebraska	2652	7667	942	11261	0.	0.	0.	0.
North Dakota	4671	8286	0	12957	0.	0.	0.	0.
Ohio	7197	0	3477	10674	0.	0.	0.	0.
Oklahoma	4030	1864	4339	10233	0.	0.	0.	0.
South Dakota	24	11013	5526	16563	0.	0.	0.	0.
Tennessee	2242	645	8325	11212	0.	0.	0.	0.
Wisconsin	13309	1280	3720	18309	0.	0.	0.	0.
PAD District III	15248	237xx*	26417	653xx*	0.	-19.A	0.	-19.A
Alabama	3504	2xx*	2856	66xx*	0.	25.w	0.	25.A
Arkansas	1598	5628	6358	13584	0.	0.	0.	0.
Louisiana	0	3088	4667	7755	0.	0.	0.	0.
Mississippi	666	8925	2980	12571	0.	0.	0.	0.
New Mexico	8410	4928	6696	20034	0.	0.	0.	0.
Texas	1070	8xx*	2860	47xx*	0.	-44.w	0.	-44.A
PAD District IV	13xxx*	23xxx*	84xx*	451xx*	-51.A	132.A	-44.A	37.A
Colorado	0	8772	5637	14409	0.	0.	0.	0.
Idaho	925	1xxx*	8xx*	2xxx*	0.	94.w	-44.w	50.A
Montana	xxx*	7358	0	7xxx*	-51.w	0.	0.	-51.A
Utah	5676	4xx*	0	61xx*	0.	38.w	0.	38.A
Wyoming	6446	5660	1952	14058	0.	0.	0.	0.
PAD District V	19060	36xxx*	14709	70xxx*	0.	-69.A	0.	-69.A
Alaska	0	7948	4300	12248	0.	0.	0.	0.
Arizona	2721	828	2189	5738	0.	0.	0.	0.
California	3792	3xxx*	2251	9xxx*	0.	-69.A	0.	-69.A
Hawaii	1038	6141	327	7506	0.	0.	0.	0.
Nevada	2555	3522	0	6077	0.	0.	0.	0.
Oregon	0	14325	3040	17365	0.	0.	0.	0.
Washington	8954	0	2602	11556	0.	0.	0.	0.

* Lower Digits Withheld. Table Cell Perturbation Error Exceeds .010%

Appendix B

Noise Addition – 0.01% Perturbation Error Acceptable

Noise Based Table 01 REGULAR

	DTW	RACK	BULK	TOTAL	< ---- Adjustments ---->			
United States	188684	218xxx*	169xxx*	576xxx*	16.	-357.	-81.	-422.
PAD District I	64xxx*	72xxx*	65xxx*	2032xx*	144.	-219.	53.	-22.
Subdistrict IA	25xxx*	28xxx*	16xxx*	71xxx*	267.	115.	-83.	300.
Connecticut	6xxx*	15xx*	17xx*	9xxx*	161.	26.	35.	222.
Maine	3xxx*	4xxx*	4xxx*	12xxx*	-150.	69.	-64.	-145.
Massachusetts	1xx*	3xxx*	0	4xxx*	9.	65.	0.	74.
New Hampshire	8xxx*	0	3xxx*	11xxx*	159.	0.	-81.	78.
Rhode Island	1xxx*	6xxx*	3xxx*	11xxx*	53.	-143.	-79.	-168.
Vermont	53xx*	12xxx*	3xxx*	21xxx*	34.	98.	106.	238.
Subdistrict IB	19xxx*	16xxx*	22xxx*	582xx*	-123.	-51.	137.	-38.
Delaware	6xxx*	23xx*	42xx*	13xxx*	-133.	-19.	-47.	-200.
District of Columbia	2xxx*	4xxx*	11xxx*	18xxx*	53.	-85.	-86.	-118.
Maryland	3xxx*	18xx*	1xxx*	6xxx*	120.	46.	110.	275.
New Jersey	6xxx*	0	13x*	6xxx*	-163.	0.	-5.	-168.
New York	0	6xx*	8xx*	14xx*	0.	-22.	41.	19.
Pennsylvania	0	65xx*	4xxx*	11xxx*	0.	30.	124.	154.
Subdistrict IC	19894	27xxx*	26332	73xxx*	0.	-283.	-1.	-284.
Florida	0	108xx*	18xx*	12703	0.	-15.	15.	-1.
Georgia	9xxx*	0	0	9xxx*	-93.	0.	0.	-93.
North Carolina	2xxx*	7xxx*	85xx*	17xxx*	-94.	-205.	49.	-250.
South Carolina	11xx*	590x*	7xxx*	14xxx*	-25.	5.	-165.	-186.
Virginia	3xxx*	0	3xxx*	7xxx*	187.	0.	112.	298.
West Virginia	29xx*	3xxx*	48xx*	11xxx*	26.	-68.	-10.	-52.
PAD District II	761xx*	61xxx*	547xx*	192xxx*	-12.	-186.	-22.	-220.
Illinois	3xxx*	0	0	3xxx*	-154.	0.	0.	-154.
Indiana	4xxx*	0	3xxx*	84xx*	-158.	0.	112.	-46.
Iowa	1xxx*	42xx*	4xxx*	9xxx*	-50.	49.	-60.	-62.
Kansas	12xxx*	10xxx*	19xx*	24xxx*	62.	156.	-21.	197.
Kentucky	5xxx*	2xxx*	6xxx*	15xxx*	-106.	-61.	73.	-94.
Michigan	1xxx*	0	6xxx*	8xxx*	-73.	0.	158.	85.
Minnesota	6xxx*	3xxx*	12xx*	1142x*	145.	-110.	-37.	-2.
Missouri	6xxx*	103xx*	39xx*	20xxx*	195.	-46.	-13.	136.
Nebraska	2xxx*	7xxx*	9xx*	11xxx*	80.	-148.	-32.	-100.
North Dakota	46xx*	83xx*	0	129xx*	-30.	18.	0.	-12.
Ohio	7xxx*	0	35xx*	10xxx*	101.	0.	40.	141.
Oklahoma	39xx*	1xxx*	4xxx*	10xxx*	-46.	52.	-82.	-76.
South Dakota	2x*	10xxx*	5xxx*	16xxx*	-1.	-75.	-94.	-170.
Tennessee	2xxx*	6xx*	8xxx*	11xxx*	70.	14.	-157.	-73.
Wisconsin	132xx*	12xx*	3xxx*	183xx*	-47.	-35.	91.	9.
PAD District III	152xx*	23xxx*	26xxx*	654xx*	-30.	233.	-181.	22.
Alabama	3xxx*	2xx*	2xxx*	6xxx*	94.	26.	51.	171.
Arkansas	15xx*	56xx*	6xxx*	135xx*	-11.	43.	-80.	-48.
Louisiana	0	31xx*	4xxx*	7xxx*	0.	21.	-115.	-94.
Mississippi	6xx*	9xxx*	3xxx*	12xxx*	23.	218.	155.	396.
New Mexico	8xxx*	4xxx*	6xxx*	19xxx*	-158.	-121.	-111.	-389.
Texas	11xx*	9xx*	2xxx*	48xx*	22.	46.	-81.	-14.
PAD District IV	13xxx*	23xxx*	8xxx*	451xx*	106.	133.	-228.	11.
Colorado	0	8xxx*	5xxx*	143xx*	0.	116.	-129.	-13.
Idaho	9xx*	xxx*	8xx*	2xxx*	19.	-100.	-49.	-131.
Montana	xxx*	73xx*	0	78xx*	-53.	23.	0.	-30.
Utah	5xxx*	4xx*	0	6xxx*	99.	39.	0.	138.
Wyoming	64xx*	5xxx*	19xx*	141xx*	41.	54.	-49.	46.
PAD District V	18xxx*	36xxx*	15xxx*	70xxx*	-192.	-318.	297.	-213.
Alaska	0	79xx*	4xxx*	12xxx*	0.	-36.	87.	51.
Arizona	2xxx*	7xx*	2xxx*	56xx*	-82.	-29.	67.	-44.
California	3xxx*	3xxx*	2xxx*	98xx*	84.	-97.	61.	48.
Hawaii	xxx*	6xxx*	3xx*	7xxx*	-76.	196.	-11.	108.
Nevada	25xx*	3xxx*	0	60xx*	19.	-56.	0.	-37.
Oregon	0	14xxx*	3xxx*	17xxx*	0.	-297.	64.	-233.
Washington	8xxx*	0	26xx*	11xxx*	-136.	0.	30.	-106.

* Lower Digits Withheld. Table Cell Perturbation Error Exceeds .010%

Appendix C

CTA Targeted Towards Smaller Tabular Cells – 0.01% Perturbation Error Acceptable

CTA Based Table 01 [TARGET SMALL CELLS] REGULAR

	DTW	RACK	BULK	TOTAL	< ---- Adjustments ---->			
United States	188668	218471	170021	577160	0.	0.	0.	0.
PAD District I	64625	72994	65620	203239	0.	0.	0.	0.
Subdistrict IA	25314	28780	16952	71046	0.	0.	0.	0.
Connecticut	62xx*	15xx*	1700	9452	-8.A	8.A	0.	0.
Maine	3936	4719	4429	13084	0.	0.	0.	0.
Massachusetts	1xx*	38xx*	0	4012	8.w	-8.A	0.	0.
New Hampshire	7879	0	3188	11067	0.	0.	0.	0.
Rhode Island	1748	6224	3976	11948	0.	0.	0.	0.
Vermont	5321	12503	3659	21483	0.	0.	0.	0.
Subdistrict IB	19417	16493	22335	58245	0.	0.	0.	0.
Delaware	6978	2xxx*	4xxx*	13650	0.	85.A	-85.A	0.
District of Columbia	2253	5070	11338	18661	0.	0.	0.	0.
Maryland	3311	1xxx*	1xxx*	6226	0.	-60.A	60.w	0.
New Jersey	6875	0	1xx*	70xx*	0.	0.	-14.A	-14.A
New York	0	6xx*	8xx*	14xx*	0.	-25.A	39.w	14.A
Pennsylvania	0	6539	4718	11257	0.	0.	0.	0.
Subdistrict IC	19894	27721	26333	73948	0.	0.	0.	0.
Florida	0	10857	1847	12704	0.	0.	0.	0.
Georgia	9961	0	0	9961	0.	0.	0.	0.
North Carolina	2268	72xx*	84xx*	17958	0.	39.A	-39.A	0.
South Carolina	1195	5887	7582	14664	0.	0.	0.	0.
Virginia	3560	0	3625	7185	0.	0.	0.	0.
West Virginia	2910	37xx*	48xx*	11476	0.	-39.A	39.A	0.
PAD District II	76174	62147	54796	193117	0.	0.	0.	0.
Illinois	4128	0	0	4128	0.	0.	0.	0.
Indiana	45xx*	0	38xx*	8459	-33.A	0.	33.A	0.
Iowa	1149	4196	4216	9561	0.	0.	0.	0.
Kansas	12xxx*	1032x*	1xxx*	24274	115.A	-3.A	-112.A	0.
Kentucky	5xxx*	280x*	6xxx*	15136	-61.A	3.A	58.A	0.
Michigan	2022	0	6668	8690	0.	0.	0.	0.
Minnesota	6400	3694	1332	11426	0.	0.	0.	0.
Missouri	5915	10385	3934	20234	0.	0.	0.	0.
Nebraska	2652	7667	942	11261	0.	0.	0.	0.
North Dakota	4671	8286	0	12957	0.	0.	0.	0.
Ohio	7197	0	3477	10674	0.	0.	0.	0.
Oklahoma	40xx*	1864	43xx*	10233	-21.A	0.	21.A	0.
South Dakota	24	11013	5526	16563	0.	0.	0.	0.
Tennessee	2242	645	8325	11212	0.	0.	0.	0.
Wisconsin	13309	1280	3720	18309	0.	0.	0.	0.
PAD District III	15248	23726	26417	65391	0.	0.	0.	0.
Alabama	35xx*	2xx*	2856	6619	25.A	-25.w	0.	0.
Arkansas	1598	5628	6358	13584	0.	0.	0.	0.
Louisiana	0	3xxx*	4xxx*	7755	0.	65.A	-65.A	0.
Mississippi	666	9xxx*	2xxx*	12571	0.	69.A	-69.A	0.
New Mexico	8410	4928	6696	20034	0.	0.	0.	0.
Texas	10xx*	xxx*	3xxx*	4828	-25.A	-109.w	134.A	0.
PAD District IV	13561	23112	8479	45152	0.	0.	0.	0.
Colorado	0	8772	5637	14409	0.	0.	0.	0.
Idaho	1xxx*	xxx*	8xx*	2xxx*	75.A	-94.w	-44.w	-63.A
Montana	xxx*	7xxx*	0	79xx*	-104.w	132.A	0.	28.A
Utah	57xx*	3xx*	0	60xx*	29.A	-38.w	0.	-9.A
Wyoming	6446	5660	20xx*	141xx*	0.	0.	44.A	44.A
PAD District V	19060	36492	14709	70261	0.	0.	0.	0.
Alaska	0	7948	4300	12248	0.	0.	0.	0.
Arizona	2721	828	2189	5738	0.	0.	0.	0.
California	3792	3728	2251	9771	0.	0.	0.	0.
Hawaii	1038	6141	327	7506	0.	0.	0.	0.
Nevada	2555	3522	0	6077	0.	0.	0.	0.
Oregon	0	14325	3040	17365	0.	0.	0.	0.
Washington	8954	0	2602	11556	0.	0.	0.	0.

* Lower Digits Withheld. Table Cell Perturbation Error Exceeds .010%

Appendix D

CTA Targeted Towards Larger Tabular Cells – 1% Perturbation Error Acceptable

CTA Based Table 01 [TARGET LARGE CELLS]   REGULAR

	DTW	RACK	BULK	TOTAL	< ---- Adjustments ---->			
United States	188538	218584	169960	577082	-130.A	113.A	-61.A	-78.A
PAD District I	64617	73063	65477	203157	-8.A	69.A	-143.A	-82.A
Subdistrict IA	25306	28788	16952	71046	-8.A	8.A	0.	0.
Connecticut	6258	1494	1700	9452	0.	0.	0.	0.
Maine	3936	4719	4429	13084	0.	0.	0.	0.
Massachusetts	1xx*	3848	0	4012	-8.w	8.A	0.	0.
New Hampshire	7879	0	3188	11067	0.	0.	0.	0.
Rhode Island	1748	6224	3976	11948	0.	0.	0.	0.
Vermont	5321	12503	3659	21483	0.	0.	0.	0.
Subdistrict IB	19417	16541	22274	58232	0.	48.A	-61.A	-13.A
Delaware	6978	24xx*	4272	13698	0.	48.A	0.	48.A
District of Columbia	2253	5070	11338	18661	0.	0.	0.	0.
Maryland	3311	1836	1xxx*	6166	0.	0.	-60.w	-60.A
New Jersey	6875	0	144	7019	0.	0.	0.	0.
New York	0	648	7xx*	13xx*	0.	0.	-39.w	-39.A
Pennsylvania	0	6539	4756	11295	0.	0.	38.A	38.A
Subdistrict IC	19894	27734	26251	73879	0.	13.A	-82.A	-69.A
Florida	0	10857	1830	12687	0.	0.	-17.A	-17.A
Georgia	9961	0	0	9961	0.	0.	0.	0.
North Carolina	2268	7239	8399	17906	0.	13.A	-65.A	-52.A
South Carolina	1195	5887	7582	14664	0.	0.	0.	0.
Virginia	3560	0	3625	7185	0.	0.	0.	0.
West Virginia	2910	3751	4815	11476	0.	0.	0.	0.
PAD District II	76103	62147	54922	193172	-71.A	0.	126.A	55.A
Illinois	4128	0	0	4128	0.	0.	0.	0.
Indiana	4599	0	3860	8459	-14.A	0.	14.A	0.
Iowa	1149	4196	4216	9561	0.	0.	0.	0.
Kansas	11939	10330	2xxx*	24329	-57.A	0.	112.A	55.A
Kentucky	5826	2787	6523	15136	0.	0.	0.	0.
Michigan	2022	0	6668	8690	0.	0.	0.	0.
Minnesota	6400	3694	1332	11426	0.	0.	0.	0.
Missouri	5915	10385	3934	20234	0.	0.	0.	0.
Nebraska	2652	7667	942	11261	0.	0.	0.	0.
North Dakota	4671	8286	0	12957	0.	0.	0.	0.
Ohio	7197	0	3477	10674	0.	0.	0.	0.
Oklahoma	4030	1864	4339	10233	0.	0.	0.	0.
South Dakota	24	11013	5526	16563	0.	0.	0.	0.
Tennessee	2242	645	8325	11212	0.	0.	0.	0.
Wisconsin	13309	1280	3720	18309	0.	0.	0.	0.
PAD District III	15248	23707	26417	65372	0.	-19.A	0.	-19.A
Alabama	3504	2xx*	2856	6644	0.	25.w	0.	25.A
Arkansas	1598	5628	6358	13584	0.	0.	0.	0.
Louisiana	0	3088	4667	7755	0.	0.	0.	0.
Mississippi	666	8925	2980	12571	0.	0.	0.	0.
New Mexico	8410	4928	6696	20034	0.	0.	0.	0.
Texas	1070	8xx*	2860	4784	0.	-44.w	0.	-44.A
PAD District IV	13510	23244	8435	45189	-51.A	132.A	-44.A	37.A
Colorado	0	8772	5637	14409	0.	0.	0.	0.
Idaho	925	1xxx*	8xx*	2xxx*	0.	94.w	-44.w	50.A
Montana	xxx*	7358	0	7821	-51.w	0.	0.	-51.A
Utah	5676	4xx*	0	6096	0.	38.w	0.	38.A
Wyoming	6446	5660	1952	14058	0.	0.	0.	0.
PAD District V	19060	36423	14709	70192	0.	-69.A	0.	-69.A
Alaska	0	7948	4300	12248	0.	0.	0.	0.
Arizona	2721	828	2189	5738	0.	0.	0.	0.
California	3792	3xxx*	2251	9702	0.	-69.A	0.	-69.A
Hawaii	1038	6141	327	7506	0.	0.	0.	0.
Nevada	2555	3522	0	6077	0.	0.	0.	0.
Oregon	0	14325	3040	17365	0.	0.	0.	0.
Washington	8954	0	2602	11556	0.	0.	0.	0.

* Lower Digits Withheld. Table Cell Perturbation Error Exceeds 1.000%

Appendix E

Noise Addition – 1% Perturbation Error Acceptable

Noise Based Table 01 REGULAR

	DTW	RACK	BULK	TOTAL	< ----	Adjustments	---->	
United States	188684	218113	169940	576738	16.	-357.	-81.	-422.
PAD District I	64769	72775	65673	203217	144.	-219.	53.	-22.
Subdistrict IA	25xxx*	28895	16869	71345	267.	115.	-83.	300.
Connecticut	6xxx*	15xx*	17xx*	9xxx*	161.	26.	35.	222.
Maine	3xxx*	4xxx*	4xxx*	12xxx*	-150.	69.	-64.	-145.
Massachusetts	1xx*	3xxx*	0	4xxx*	9.	65.	0.	74.
New Hampshire	8xxx*	0	3xxx*	11145	159.	0.	-81.	78.
Rhode Island	1xxx*	6xxx*	3xxx*	11xxx*	53.	-143.	-79.	-168.
Vermont	5355	12600	3xxx*	21xxx*	34.	98.	106.	238.
Subdistrict IB	19293	16442	22471	58207	-123.	-51.	137.	-38.
Delaware	6xxx*	2380	42xx*	13xxx*	-133.	-19.	-47.	-200.
District of Columbia	2xxx*	4xxx*	11251	18543	53.	-85.	-86.	-118.
Maryland	3xxx*	18xx*	1xxx*	6xxx*	120.	46.	110.	275.
New Jersey	6xxx*	0	13x*	6xxx*	-163.	0.	-5.	-168.
New York	0	6xx*	8xx*	14xx*	0.	-22.	41.	19.
Pennsylvania	0	6568	4xxx*	11xxx*	0.	30.	124.	154.
Subdistrict IC	19894	27xxx*	26332	73664	0.	-283.	-1.	-284.
Florida	0	10841	1861	12703	0.	-15.	15.	-1.
Georgia	9868	0	0	9868	-93.	0.	0.	-93.
North Carolina	2xxx*	7xxx*	8512	17xxx*	-94.	-205.	49.	-250.
South Carolina	11xx*	5891	7xxx*	14xxx*	-25.	5.	-165.	-186.
Virginia	3xxx*	0	3xxx*	7xxx*	187.	0.	112.	298.
West Virginia	2936	3xxx*	4804	11424	26.	-68.	-10.	-52.
PAD District II	76161	61960	54774	192896	-12.	-186.	-22.	-220.
Illinois	3xxx*	0	0	3xxx*	-154.	0.	0.	-154.
Indiana	4xxx*	0	3xxx*	8412	-158.	0.	112.	-46.
Iowa	1xxx*	42xx*	4xxx*	9499	-50.	49.	-60.	-62.
Kansas	12057	10xxx*	19xx*	24470	62.	156.	-21.	197.
Kentucky	5xxx*	2xxx*	6xxx*	15042	-106.	-61.	73.	-94.
Michigan	1xxx*	0	6xxx*	8775	-73.	0.	158.	85.
Minnesota	6xxx*	3xxx*	12xx*	11424	145.	-110.	-37.	-2.
Missouri	6xxx*	10339	3920	20369	195.	-46.	-13.	136.
Nebraska	2xxx*	7xxx*	9xx*	11161	80.	-148.	-32.	-100.
North Dakota	4640	8304	0	12945	-30.	18.	0.	-12.
Ohio	7xxx*	0	35xx*	10xxx*	101.	0.	40.	141.
Oklahoma	39xx*	1xxx*	4xxx*	10156	-46.	52.	-82.	-76.
South Dakota	2x*	10938	5xxx*	16xxx*	-1.	-75.	-94.	-170.
Tennessee	2xxx*	6xx*	8xxx*	11138	70.	14.	-157.	-73.
Wisconsin	13261	12xx*	3xxx*	18317	-47.	-35.	91.	9.
PAD District III	15217	23959	26235	65413	-30.	233.	-181.	22.
Alabama	3xxx*	2xx*	2xxx*	6xxx*	94.	26.	51.	171.
Arkansas	1586	5671	6xxx*	13535	-11.	43.	-80.	-48.
Louisiana	0	3109	4xxx*	7xxx*	0.	21.	-115.	-94.
Mississippi	6xx*	9xxx*	3xxx*	12xxx*	23.	218.	155.	396.
New Mexico	8xxx*	4xxx*	6xxx*	19xxx*	-158.	-121.	-111.	-389.
Texas	11xx*	9xx*	2xxx*	4814	22.	46.	-81.	-14.
PAD District IV	13667	23244	8xxx*	45163	106.	133.	-228.	11.
Colorado	0	8xxx*	5xxx*	14395	0.	116.	-129.	-13.
Idaho	9xx*	xxx*	8xx*	2xxx*	19.	-100.	-49.	-131.
Montana	xxx*	7381	0	7842	-53.	23.	0.	-30.
Utah	5xxx*	4xx*	0	6xxx*	99.	39.	0.	138.
Wyoming	6487	5714	19xx*	14104	41.	54.	-49.	46.
PAD District V	18xxx*	36173	15xxx*	70047	-192.	-318.	297.	-213.
Alaska	0	7911	4xxx*	12298	0.	-36.	87.	51.
Arizona	2xxx*	7xx*	2xxx*	5694	-82.	-29.	67.	-44.
California	3xxx*	3xxx*	2xxx*	9818	84.	-97.	61.	48.
Hawaii	xxx*	6xxx*	3xx*	7xxx*	-76.	196.	-11.	108.
Nevada	2573	3xxx*	0	6039	19.	-56.	0.	-37.
Oregon	0	14xxx*	3xxx*	17xxx*	0.	-297.	64.	-233.
Washington	8xxx*	0	26xx*	11449	-136.	0.	30.	-106.

* Lower Digits Withheld. Table Cell Perturbation Error Exceeds 1.000%

Appendix F

CTA Targeted Towards Smaller Tabular Cells – 1% Perturbation Error Acceptable

CTA Based Table 01 [TARGET SMALL CELLS] REGULAR

	DTW	RACK	BULK	TOTAL	< ---- Adjustments ---->			
United States	188668	218471	170021	577160	0.	0.	0.	0.
PAD District I	64625	72994	65620	203239	0.	0.	0.	0.
Subdistrict IA	25314	28780	16952	71046	0.	0.	0.	0.
Connecticut	6250	1502	1700	9452	-8.A	8.A	0.	0.
Maine	3936	4719	4429	13084	0.	0.	0.	0.
Massachusetts	1xx*	3832	0	4012	8.w	-8.A	0.	0.
New Hampshire	7879	0	3188	11067	0.	0.	0.	0.
Rhode Island	1748	6224	3976	11948	0.	0.	0.	0.
Vermont	5321	12503	3659	21483	0.	0.	0.	0.
Subdistrict IB	19417	16493	22335	58245	0.	0.	0.	0.
Delaware	6978	2xxx*	4xxx*	13650	0.	85.A	-85.A	0.
District of Columbia	2253	5070	11338	18661	0.	0.	0.	0.
Maryland	3311	1xxx*	1xxx*	6226	0.	-60.A	60.w	0.
New Jersey	6875	0	1xx*	7005	0.	0.	-14.A	-14.A
New York	0	6xx*	8xx*	1446	0.	-25.A	39.w	14.A
Pennsylvania	0	6539	4718	11257	0.	0.	0.	0.
Subdistrict IC	19894	27721	26333	73948	0.	0.	0.	0.
Florida	0	10857	1847	12704	0.	0.	0.	0.
Georgia	9961	0	0	9961	0.	0.	0.	0.
North Carolina	2268	7265	8425	17958	0.	39.A	-39.A	0.
South Carolina	1195	5887	7582	14664	0.	0.	0.	0.
Virginia	3560	0	3625	7185	0.	0.	0.	0.
West Virginia	2910	37xx*	4854	11476	0.	-39.A	39.A	0.
PAD District II	76174	62147	54796	193117	0.	0.	0.	0.
Illinois	4128	0	0	4128	0.	0.	0.	0.
Indiana	4580	0	3879	8459	-33.A	0.	33.A	0.
Iowa	1149	4196	4216	9561	0.	0.	0.	0.
Kansas	12111	10327	1xxx*	24274	115.A	-3.A	-112.A	0.
Kentucky	5xxx*	2790	6581	15136	-61.A	3.A	58.A	0.
Michigan	2022	0	6668	8690	0.	0.	0.	0.
Minnesota	6400	3694	1332	11426	0.	0.	0.	0.
Missouri	5915	10385	3934	20234	0.	0.	0.	0.
Nebraska	2652	7667	942	11261	0.	0.	0.	0.
North Dakota	4671	8286	0	12957	0.	0.	0.	0.
Ohio	7197	0	3477	10674	0.	0.	0.	0.
Oklahoma	4009	1864	4360	10233	-21.A	0.	21.A	0.
South Dakota	24	11013	5526	16563	0.	0.	0.	0.
Tennessee	2242	645	8325	11212	0.	0.	0.	0.
Wisconsin	13309	1280	3720	18309	0.	0.	0.	0.
PAD District III	15248	23726	26417	65391	0.	0.	0.	0.
Alabama	3529	2xx*	2856	6619	25.A	-25.w	0.	0.
Arkansas	1598	5628	6358	13584	0.	0.	0.	0.
Louisiana	0	3xxx*	4xxx*	7755	0.	65.A	-65.A	0.
Mississippi	666	8994	2xxx*	12571	0.	69.A	-69.A	0.
New Mexico	8410	4928	6696	20034	0.	0.	0.	0.
Texas	10xx*	xxx*	3xxx*	4828	-25.A	-109.w	134.A	0.
PAD District IV	13561	23112	8479	45152	0.	0.	0.	0.
Colorado	0	8772	5637	14409	0.	0.	0.	0.
Idaho	1xxx*	xxx*	8xx*	2xxx*	75.A	-94.w	-44.w	-63.A
Montana	xxx*	7xxx*	0	7900	-104.w	132.A	0.	28.A
Utah	5705	3xx*	0	6049	29.A	-38.w	0.	-9.A
Wyoming	6446	5660	20xx*	14102	0.	0.	44.A	44.A
PAD District V	19060	36492	14709	70261	0.	0.	0.	0.
Alaska	0	7948	4300	12248	0.	0.	0.	0.
Arizona	2721	828	2189	5738	0.	0.	0.	0.
California	3792	3728	2251	9771	0.	0.	0.	0.
Hawaii	1038	6141	327	7506	0.	0.	0.	0.
Nevada	2555	3522	0	6077	0.	0.	0.	0.
Oregon	0	14325	3040	17365	0.	0.	0.	0.
Washington	8954	0	2602	11556	0.	0.	0.	0.

* Lower Digits Withheld. Table Cell Perturbation Error Exceeds 1.000%

Comments by Reviewers

***************************** Review 1 *********************
Title: How to Prevent Cell Perturbation Procedures from Becoming Data Falsification Procedures (Incorporating Quality Measures in Tabular Data Protected by Cell Perturbation Methods) A Practical Solution

Authors: Ramesh A. Dandekar
Overall Evaluation: 3
Confidence: 3

Verbal Evaluation: The paper is in the context of perturbative methods for tabular data protection such as controlled tabular adjustment or micro data level noise addition proposed and/or discussed in earlier papers of this author. The author proposes in this paper to suppress the right most digits of cell values that have been perturbed beyond a pre-defined threshold. This is an interesting and perhaps relavant proposition, although rather trivial, and neither theoretically nor empirically supported in the paper. The main contribution of the paper is to demonstrate the performance of the suggestion under different scenarios on synthetic tables with real life structures.

The presentation is clear enough, but the empirical demonstration is too dominant in the paper. The title is rather an abstract than a title. With one exception, no reference to the work of others in this field is given.

Comments to Authors: You should include more motivation for your formula to determine the number of right most digits to suppress. You claim that your formula "eliminates the possibility of indirect disclosure ...by...external pattern auditing procudures", but you do not support this claim by any kind of "proof". Suppressing the right most digits means that an interval for the adjusted value is published. Does the interval always include the true original value? In case it does, it is also an interval for the true value. What, if the intervals are used by external auditing procedures? Considering this aspect, you should discuss the question, if the intervals provide upper and lower, or only sliding protection for sensitive cells. You might also consider to compare your method to other interval publication methods like partial suppression.

Comments to PC:

***************************** Review 2 *********************

Title: How to Prevent Cell Perturbation Procedures from Becoming Data Falsification Procedures
Authors: Dandekar
Overall Evaluation: 7
Confidence: 3

Verbal Evaluation:

Comments to Authors:

- Why factor of "2" in the formula for suppressed digits. Why not 1.5 or 3.Did you perform some simulations to get 2 as the right value?

- Your approach, indirectly, provides intervals, and it is related to interval protection. Can you comment on this?

- How did you choose the protection senses in CTA: heuristically or solving the difficult mixed integer-linear problem?

- Which objective was used for targeting large or small cells in CTA?

- Page 3. Don't write full title of references in the text, just cite it.

- Whenever possible, I would suggest summarizing the output from tables 2-6 and those in the appendix in a more effective way.

Comments to PC:

***************************** Review 3 *********************
Paper #07:
Title: How to Prevent Cell Perturbation Procedures from Becoming Data Falsification Procedures
Authors: Ramesh Dandekar
Overall Evaluation: 3

Confidence: 3

Verbal Evaluation:

The paper was weak, disorganized and I don't think the topic has much merit. It suggests combining a cell suppression strategy with cell perturbation and I do not see the point of doing this. The paper also covers other topics which are not specifically stated in the introduction, such as different target functions for the cell perturbation method, comparison of cell perturbation with adding noise in the micro data. These topics should have been discussed in the introduction or as customary the different sections of the paper should have been described in the introduction to let readers know the content of the paper. I would have liked to have seen some referencing to other work besides the author's own references. More details are presented in the comments to the author but in general I do not think this paper should be included in the conference proceedings. The positive aspect of the paper is the use of a quality indicator to let users know the extent of the perturbation to the cells.

Comments to Authors:

The introduction to your paper does not sufficiently describe the contents of the paper. Some of the topics that I found throughout your paper were:
1. Providing a quality indicator to users on the amount of perturbation
2. Combining a cell suppression strategy with cell perturbation based on CTA
2. Comparing different target functions in the CTA approach
3. Comparison of CTA to adding noise in the weights of the micro data

These topics should have been defined in the introduction with appropriate references to the different sections of the paper. In addition, there is almost no referencing to other work carried out in this area and aside from one reference, all the rest of the references pertain to your work. This narrows the scope of the paper. The abstract claims that you will be able to "convey the relative standard errors associated with tabular format estimates derived from sampled survey data". This is usually carried out in general for sampled survey data (regardless of SDL methods) and in any case, it is not clear where this is discussed in the paper itself.

I found it difficult to understand the motivation of your idea of combining a cell suppression technique with a cell perturbation method. Your examples in the introduction of looking at trends in the data would certainly have to be accompanied by confidence intervals in order to determine significance. SDL methods are generally kept to within the limits of sampling error and for more sophisticated methods ensure that sampling errors before and after the perturbation are preserved. It is not clear to me why a Data Supplier would want to carry out both cell suppression and a tabular cell perturbation method which greatly increases loss of information. In fact, by punching so many holes in the data, users won't be able to look at trends at all in the data. In addition, compared to standard cell suppression strategies based on high risk cells (low frequency count or dominance in the cell), you seem to suggest suppressing large cells as well.

The cell suppression technique is described as a "quality preserving" solution, although it actually causes much more information loss. The formula for determining the number of digits to suppress needs more explanation. The comparison of methods as described in the paper is not coherent. You provide, for example Table 3 and Table 4 for comparisons but it is difficult to note the differences and if they are significant. Better presentation of your results should have been carried out. In addition, as in all post-tabular methods, the problem of consistency across tables needs to be addressed. The idea of presenting a threshold quality indicator is very useful and a good step in providing users with knowledge about the SDL methods to take these into account when carrying out analysis.

Based on these comments, I do not recommend including the paper in the Conference proceedings.

*
Comments to PC: none
*