General Discrete-data Modeling Methods for Producing Synthetic Data with Reduced Re-identification Risk that Preserve Analytic Properties

William E. Winkler[1], william.e.winkler@census.gov 2008Nov12
U.S. Census Bureau, Statistical Research Division, Washington, DC 20233-9100

Abstract

General modeling methods for representing and improving the quality of discrete data (Winkler 2003, 2008) extend and connect the editing methods of Fellegi and Holt (1976) and the imputation ideas of Little and Rubin (2002). This paper describes a modeling framework to produce synthetic microdata that better corresponds to external benchmark constraints on certain aggregates (such as margins) and on which certain cell probabilities are bounded both below and above to reduce re-identification risk. Rather than use linear constraints (Meng and Rubin 1993), the modeling methods use convex constraints (Winkler 1990, 1993) in an extended MCECM procedure.

1. Introduction

This paper describes modeling methods for discrete data. The methods are closely related to general modeling/edit/imputation methods (Winkler 2008) in which models can easily be created using very fast, parameter-driven software. The methods and generalized software are suitable for a wide range of discrete data. The models are used in generalized production edit/imputation software that assure that the 'corrected' data satisfy both edit restraints and preserve joint distributions in a principled manner. Furthermore, the modeling methods use convex constraints (Winkler 1993, 1990) in an EMH algorithm that generalize the linear constraints of the MCECM algorithm of Meng and Rubin (1993). An advantage of the new modeling methods is that the microdata created via the methods can have aggregates that are adjusted to certain benchmark totals.

General convex constraints provide great flexibility in creating models that approximately preserve analytic properties and reduce the re-identification risk in synthetic microdata that are created from the models. Convex constraints allow putting lower and upper bounds on individual cells or on groups of cells. In earlier work, Winkler (2007) showed how to use more elementary methods to reduce re-identification risk by putting lower and upper bounds on both small cells and sampling zeros while still approximately preserving most aggregates needed for loglinear modeling and important joint and conditional probabilities. At that time, Winkler (2007) felt that the risk of re-identification via record linkage experiments was greatly reduced in comparison to data from some previous synthetic-data-generation methods.

Epsilon-privacy represents a gold standard in terms of preventing leakage of information and in preserving privacy. Much research is needed to justify analytic properties of epsilon-private data. Dwork, McSherry, and Talwar (2007b, first two paragraphs of section 5) provide an example from 'census' data in which the amount of noise added to a table having on the order of 1,000,000 cells must be on the order to 1,000,000 (plus or minus) in each cell. In this situation and most others where rigorous epsilon-privacy has been applied, it is not clear that the resultant 'protected' microdata will meet analytic standards acceptable to most economists and statisticians.

Additionally, Xaio and Tao (2008) raise serious concerns by demonstrating that it is impractical to verify epsilon-privacy in most situations. Specifically, they prove that L^1-sensitivity of functions (Dwork et al. 2006) is NP-Hard computationally. Dwork et al. (2006) showed that computing the L^1-sensitivity of functions was needed to verify epsilon-privacy in most situations.

The notable exception to the lack of suitable analytic properties is work by Machanavajjhala et al. (2008) that preserves an extended type of epsilon-delta privacy in a very narrowly analytically focused 'on-the-map' application. Machanavajjhala et al. applied clever theoretical techniques and introduced exceptionally complex computational methods that may not be suitable for most general situations.

In this paper, we slightly extend the methods of Winkler (2008, 2007) in a manner that creates a model with a desired set of properties. To do this we place a few pairs of upper and lower bounds on key aggregates needed for the loglinear modeling while placing upper bounds and lower bounds on a very large set of small cells and sampling zeros. The idea is to *target preservation of analytic properties* in the creation of the model. To produce synthetic data, we merely randomly draw from the model in the appropriate fashion. Typically, this means almost exactly preserving the probabilities associated with originally larger cells. Most small cells in the original data are replaced by sets of sampling zeros that have positive probability in the model and that approximately preserve the key aggregates needed for loglinear modeling.

There are several key points of the new methods. First, any direct re-identification experiment will only match originally small cells with sampling zeros that have very small positive probability in the models. Second, because we target preservation of a few analytic properties, we are not creating all of the key aggregates (functions) in a manner where each function satisfies epsilon-privacy. We do create an alternative to a type of epsilon-delta-privacy that we believe would make it exceptionally difficult to reconstruct the original private data in manners suggested by Dwork (2006), Barak et al. (2007), Dwork et al. (2007a) and Dwork and Yekhanin (2008)

Although the computational algorithms needed for creating the models are sufficiently fast for the largest edit/imputation applications, the algorithms need speeding up for even moderate size (50 million cells) modeling situations needed for producing synthetic data.

In the second section of this paper, we give cursory background on edit/imputation and some of the basic computational algorithms. We also describe how a re-identification experiment is performed that assures that private data cannot be easily re-identified but may not satisfy reasonable epsilon-privacy or epsilon-delta privacy. We describe how the models are created. In the third section, we provide empirical results on 'census' data that has be downloaded from the UCI machine learning repository and used in some confidentiality research. Although any synthetic data produced from the model can prevent most re-identification using record linkage and satisfies a condition that can be considered an alternative to very weakened-type of epsilon-delta-type of privacy, the synthetic data do not satisfy rigorous epsilon-delta privacy. An interesting experiment (beyond the scope of the present paper) would be for a cryptographer to apply some of the constructive methods (e.g., Dwork 2006, Barak et al. 2007, Dwork et al. (2007a), Dwork and Yekhanin 2008) to the synthetic data to reconstruct a reasonable approximation of the original private data. The final sections consist of brief discussion and concluding remarks. This experiment would be needed regardless of the type of auxiliary information (Ganta et al. 2008) that might be available to an adversary.

2. Background

In this section we provide background on modeling/edit/imputation, need for computational speed, re-identification using record linkage, and the general iterative fitting algorithm for creating the model.

2.1 Modeling/Edit/Imputation

Modern methods for edit/imputation began with the seminal paper of Fellegi and Holt (1976, hereafter FH). With discrete data, an edit might be that a child of less than 16 could not be married. Their paper provided three principles: (1) The minimum number of fields in each edit-failing record r_0 should be changed to create an edit-passing record r_1 (*error localization*), (2) Imputation rules should be derived automatically from edit rules, and (3) When imputation is necessary, it should maintain marginal and joint distributions of fields.

The FH paper was the first to provide a method that assured that an edit-failing record r_0 could be changed into an edit-passing record r_1. To assure correct error localization, FH showed that implicit edits were needed. Implicit edits are those that can be logically derived from explicitly defined edits. Winkler (1997) provided set-covering algorithms that delineated the implicit edits, the set of which can be considered structural zeros for loglinear modeling. Although a number of statistical agencies have implemented generalized FH production systems that assure the edit-failing records can be 'corrected' to edit-passing records, none have provided FH methods that assure that the records also satisfy joint distributional characteristics from a model. The FH suggestion that hot-deck could be used for (2) and (3) is not possible due to serious deficiencies in hot-deck that were not understood when the FH paper was written (Winkler 2008).

Winkler (2003) provided the theory connecting the edits of FH with the generalized imputation of Little and Rubin (2002). An initial routine (Winkler 1997) finds the set of implicit edits (structural zeros) in a manner that is 100 times as fast as the previous fastest algorithms of Garfinkel, Kunnathur, and Liepins (1986) used by IBM in creating a large system for ISTAT (Barcaroli and Venturi. 1997). A second routine (Winkler 2008) does standard loglinear modeling under a combination of linear and convex constraints in the presence of structural zeros. In the edit setting, the iterative fitting algorithm is a type of EM algorithm as in Little and Rubin (2002). The key aspect of the second routine is having computational algorithms that are sufficiently fast for all of the survey data situations in the statistical agencies. The final routine does the error localization (Winkler 1997) using either branch-and-bound or a greedy algorithm and then fills in missing or 'to-be-changed' values according to the model (contingency table) determined by the second routine. All records are guaranteed to satisfy edits and the overall set of records preserve the probability distributions of the model.

2.2 The EMH algorithm

The general iterative fitting algorithm is extended to an EMH algorithm (Winkler 1993, 1990) for convex constraints that allow putting upper bounds on cells or convex combinations of cells. Because the set of probabilities must add to one, lower bounds can also be put on cell probabilities or simple sums of cell probabilities that might correspond to a marginal constraint. The general EMH algorithm has been used for unsupervised learning of optimal record linkage parameters (Winkler 1993) in which certain probabilities are estimated within restricted ranges based on a priori knowledge. The general EMH algorithm has also been used in statistical matching to create microdata that better corresponds to (external) benchmark constraints (D'Orazio et al. 2006).

In the application of this paper, we apply the EMH algorithm with several constraints. First, we perform standard loglinear modeling to determine the set of interactions needed to get suitably close-fitting model. The model is the final set of probabilities associated with the cells corresponding to the entire set of data patterns. Second, we take the set of counts associated with the small cells (here either 1 or 2) and disperse all of the counts across the entire set of small cells and the entire set of sampling zeros. The intent is to assure positive probability of sampling zeros in a manner that preserves most of the characteristics of the best-fitting set of interactions under purely linear constraints. Third, we place upper bounds (say 0.000004) on the probabilities associated with the originally small cells that assure that the final fitted probabilities are zero to five decimal places. Fourth, if necessary, we can place upper and lower bounds on a few of the marginal probabilities in the final fitted contingency table that deviate substantially from the marginal probabilities in the original, private data.

To create the synthetic data, we randomly draw from the contingency table probability proportional to size. If necessary, we can create multiple copies of the synthetic data.

2.3 Re-identification via Record Linkage

After modeling and creation of synthetic data **Y** from the original data **X**, we can perform re-identification experiments. To do this we merely match data **Y** directly against data **X**. The re-identification experiment is conservative in the sense that it that any intruder would likely have data **Y1** that is more difficult to match against **X** than **Y**. In a real-world situation, the intruder would have names and other identifying information associated with individual records in data **Y1**. Based on the worst-case re-identifications, it is possible to extrapolate downward explicit re-identifications of individual records or of overall re-identification rates. The downward extrapolation can be based on assumed typographical error rates or the record linkage metrics that are used to compare individual fields. With discrete data, we might only do exact comparison of individual fields and use an EM-latent class algorithm for estimating the best record linkage parameters. Kim and Winkler (1995) and Winkler (1998) used the EM algorithm and different field-comparison metrics for re-identification with continuous data. For convenience, we assume that we are using entire populations so that we need not extrapolate for different sampling scenarios.

Any record corresponding to a small cell in the data **Y** that can be associated via record linkage with the correct corresponding cell in **X** with high matching probability can be considered a re-identification. With continuous data scenarios, both Fuller (1993) and Winkler (1998) showed how to perform the matching to get explicit re-identification. Discrete-data re-identification is much more straightforward under the complete population scenario of this paper. Typically, if we randomly draw synthetic data from the model of section 2.2, *we will not get any re-identification* using record linkage. The key issue with the synthetic data is whether the synthetic data preserves a few analytic constraints so that someone using the synthetic data **Y** would approximately reproduce results that could be obtained from the original data **X**.

With epsilon-privacy (e.g., Dwork 2006), individuals make similar assumptions about the best possible data **Y1** (or **Y**) that might be matched against data **X**. Epsilon-privacy goes further in that it assures almost no leakage of information that prevents re-identification but does not presently preserve analytic properties in any clearly established manner. Ganta et al. (2008) explicitly bring in the use of auxiliary information in demonstrating that epsilon-privacy prevents any type of re-identification.

2.4 The Empirical Data and Restraints Used for Modeling

Data are from the University of California at Irvine machine learning respository. The specific data set is 'Adult'. The variables (fields) downloaded were age, WorkClass (8 values), Education (16 values), MaritalStatus (7 values), Occupation (14 values), Race (5 values), Sex (2 values), and Country (41 values). For initial testing purposes, we used WorkClass (7 values), MaritalStatus (7 values), Race (5 values), and Sex (2 values) that yielded 490 ($7 \times 7 \times 5 \times 2$) data patterns. There are 45221 data records and there are no missing fields within data records. WorkClass is reduced to 7 values because one of its values (NoWork) never occurs in the data set.

The data have 80 small cells having count 1 or 2, 191 cells that are sampling zeros, and 290 cells having count above 2. The total count associated with the small cells is 103. We determine that the all 3-way interaction model gives good fits with linear constraints only. We use an EM fitting procedure in which we disperse the total count of 103 associated with the small cells across all 271 (80 + 191) cells having small or zero counts. The starting value is 103/271 in each cell and the expected E-values are based on the current set of the parameters from the M-step. The counts of the larger cells are not varied in the modeling because we are assuming that we will not be able to effectively re-identify individual large cells in synthetic data **Y** randomly drawn from the model with the individual large cells in data **X**. After the initial fitting under linear restraints, we repeat the fitting were we place additional convex constraints (upper bounds of 0.000004) on the small cells. The synthetic data is created reproducing the counts of the non-small cells and randomly sampling from the remaining cells (both small and sampling zeros) with a probability proportional to size procedure until we achieve synthetic data **Y** of size 45221.

In earlier work, Winkler (2007) showed that the fitting and modeling methods had great flexibility in a small situation representing 48 ($4 \times 3 \times 4$) cells where nearly half of the cells were structural zeros. In more recent work, Winkler (2008) showed that the modeling methods had somewhat greater flexibility in a situation with 96 ($4 \times 3 \times 4 \times 2$) cells. The point is that, with the smallest situations, we have very little flexibility in the modeling to preserve the analytic properties. With more cells (490 data patterns), we have considerably greater flexibility in preserving analytic properties. With an even greater number of cells ($580{,}160 = 74 \times 7 \times 7 \times 16 \times 5 \times 2$), we have even greater flexibility in preserving analytic properties but may encounter computational issues (10 minutes for the general fitting procedure to converge).

3. Results

The results presented in this section are intended to represent a small situation (490 cells or data patterns) that is still quite cumbersome to present because of the large size of the tables. We present the 490-cell situation because we believe that it is adequate for illustrating how analytic properties are preserved while significantly reducing re-identification risk.

Fitting the 3-way interaction model **M1** (with linear but no convex constraints), we have that the maximum possible likelihood is -3.234682 and that the likelihood that we achieve is -3.234982. The maximum deviation allowed by the fitting software is 0.0000000000100. If we fit with the same interaction restraints and an additional restraint with an upper bound of 0.000004 on each originally small cell (model **M2**), we get the likelihood of -3.241030 that indicates a reasonably good overall fit. As our fitting uses all 3-way interactions, we need to examine how closely the 3-way margins from the limiting solution under model M2 agree with the 3-way margins from the original data. In indexing cells, we use a lexicographic ordering in which (0,0,0,0)=0, (0.0,0,1)=1, …, (6,6,4,1)=489. We obtain this with the mapping (a1, a2, a3, 4)=a1*24+a2*8+a3*2+a4*1. If $\mathbf{X}_i$, $1 \le i \le 4$, is the i^{th} variable, then $\{X_1=i_1, X_2=i_2, X_3=i_3, X_4=i_4\} = (i_1, i_2, i_3, i_4)$.

Table 1 represents original and fitted probabilities associated with a few selected individual cells. It is an excerpt from the full Table A.1 given in the Appendix. A cell with a count of 1 has probability 0.00002 and a cell with count of 2 has probability 0.00004. All of the probabilities in the table are rounded to five digits. Cells 0000-0007 show that the individual cell probabilities are reasonably close to each other. Cells 0020, 0021, and 0301 have the largest deviations. Cell 0107 is an original cell with count 1 that is given a fitted probability above zero and below 0.000004. Cells 0485-0489 are sampling zeros that are given a positive probability of approximately 0.00001. When we randomly sample from Table A.1, we have positive probability of sampling each cell but originally small cells will seldom appear in the set of synthetic records. All of the greatest deviations are associated with cells that have total probability of less than 0.003. The greatest multiplicative deviation in the remaining cells is well less than 1.0. The key issue is how well are the margins preserved.

Table 1. Original and Fitted Probabilities for Selected Cells

Cell		Original	Fitted
0000	0 0 0 0	0.02859	0.02876
0001	0 0 0 1	0.25344	0.25328
0002	0 0 1 0	0.00172	0.00163
0003	0 0 1 1	0.00781	0.00790
0004	0 0 2 0	0.00031	0.00037
0005	0 0 2 1	0.00181	0.00175
0006	0 0 3 0	0.00042	0.00042
0007	0 0 3 1	0.00210	0.00210
0020	0 2 0 0	0.09670	0.09636
0021	0 2 0 1	0.12426	0.12460
0107	1 3 3 1	0.00002	0.00000
0301	4 2 0 1	0.00637	0.00610
0485	6 6 2 1	0.00000	0.00001
0486	6 6 3 0	0.00000	0.00001

0487	6 6 3 1	0.00000	0.00001
0488	6 6 4 0	0.00000	0.00001
0489	6 6 4 1	0.00000	0.00001

Table 2 contains a few selected marginal probabilities for variables 1, 3, and 4. The largest deviations 0.000210, 0.00105, and 0.000100 occurred at marginal cells 0067, 0014, and 0054, respectively. No other specific marginal probabilities for the other interaction patterns were this large. We also give the first eight marginal probabilities. Examination of table A.2 indicates that most marginal probabilities from the fitted data are very close to the marginal probabilities from the original data. The closeness of the marginal probabilities indicates that association-rule mining and other elementary analyses of the joint and conditional probabilities should yield results from synthetic data created from Table A.1 that agree somewhat with comparable results from the original confidential data.

Table 2. Original and Fitted 3-way Margins
for Selected Marginal Cells

Pattern = 3, Variables 1,3,4		
00000	0.205988	0.205988
00001	0.427102	0.427102
00002	0.007607	0.007589
00003	0.013511	0.013518
00004	0.002211	0.002223
00005	0.003936	0.003925
00006	0.002410	0.002423
00007	0.004179	0.004146
00014	0.000133	0.000028
00054	0.000199	0.000099
00067	0.000000	0.000210

4. Discussion

Re-identification experiments may not be effective in proving the privacy of synthetic data produced according to the methods of this paper. The synthetic data do not appear to satisfy any rigorous type of epsilon- or epsilon-delta privacy. If a cryptographer were to reconstruct a moderate subset of the originally-private microdata from the synthetic data, then the reconstruction should prove that re-identification experiments are not valid in verifying the privacy of synthetic microdata in most situations.

Any reconstruction of the original data from the synthetic data would be computationally challenging in moderate size situations. In the 6-variable scenario, there are 588,160 data patterns, 9447 cells having counts of 1 or 2, and 3098 cells having counts of greater than 2. The total from all the cells is 45221. Because there are so many structural zeros (~98% of 588,160 possible cells), we have great flexibility in assigning positive probabilities to the sampling zero cells in a manner in which analytic properties are approximately preserved (much better than with the 490-cell example of this paper). After the random sampling, we have a synthetic data set (or multiple synthetic data sets) in which the small counts from 9447 cells in the original private data are placed in a suitable set of sampling zero cells.

More research needs to be done to on what it means to preserve analytic properties. In particular, there needs to be more agreement among researchers on what it means to preserve analytic properties. This paper merely shows that the overall fit of the data and almost all of the 3-way margins having larger probability agree quite closely between the fitted and original data.

The computational algorithms need to be speeded up and altered. In testing on the larger data (588,160 cells), the fitting with both linear and a very simplified set of convex constraints needed 10 minutes CPU time. With a very large set of convex constraints and a variant of the current set of algorithms for the convex constraints, the fitting takes 10-100 times as long.

5. Concluding Remarks

This paper provides methods for modeling discrete data that generalize standard loglinear modeling to methods that also include convex constraints. When properly applied, the convex constraints allow significantly reduced chance of re-identification using record linkage methods. The synthetic data randomly drawn from the models approximately (but very closely) preserve a few analytic characteristics whereas epsilon-privacy methods (Dwork et al. 2007b, first two paragraphs of section 5) have not been demonstrated to preserve analytic properties. The synthetic data created by the methods of this paper do not necessarily satisfy epsilon-privacy or epsilon-delta-privacy (Machanavajjhala et al. 2008) but might be exceptionally difficult to re-identify using cryptographic protocols and exceptionally large amounts of computation.

1/ This report is released to inform interested parties of (ongoing) research and to encourage discussion (of work in progress). Any views expressed on (statistical, methodological, technical, or operational) issues are those of the author(s) and not necessarily those of the U.S. Census Bureau.

References

Abowd, J. M. (2008), "How Protective are Synthetic Data," *Privacy in Statistical Databases 2008*, New York: Springer.

Agresti, A. (2007), *An Introduction to Categorical Data Analysis (2nd Edition)*, New York: J. Wiley.

Barak, B., Chaudhuri, K., Dwork, C., Kale, S., McSherry, F., and Talwar, K. (2007), "Privacy, Accuracy, and Consistency Too: A Holistic Solution to Contingency Table Release," PODS '07, Beijing, China.

Barcaroli, G., and Venturi, M. (1997), "DAISY (Design, Analysis and Imputation System): Structure, Methodology, and First Applications," in (J. Kovar and L. Granquist, eds.) *Statistical Data Editing, Volume II*, U.N. Economic Commission for Europe, 40-51.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W., (1975), *Discrete Multivariate Analysis*, Cambridge, MA: MIT Press.

D'Orazio, M., Di Zio, M., and Scanu, M. (2006), "Statistical Matching for Categorical Data: Displaying Uncertainty and Using Logical Constraints," *Journal of Official Statistics*, 22 (1), 137-157.

Dwork, C. (2006), "Differential Privacy," 33rd International Colloquium on Automata, Languages and Programming – ICALP 2006, Part II, 1-12.

Dwork, C. (2008), "Differential Privacy: A Survey of Results," in (M. Agrawal et al., eds.) TAMC 2008, LNCS 4978, 1-19.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006), "Calibrating Noise to Sensitivity in Private Data Analysis," 3rd Conference on Cryptography – TCC 2006, 365-384.

Dwork, C., McSherry, F., and Talwar, K. (2007a), “The Price of Privacy and the Limits of LP Decoding,” STOC ’07, San Diego, CA.
Dwork, C., McSherry, F., and Talwar, K. (2007b), “Differentially Private Marginals Release with Mutual Consistency and Error Independent Sample Size,” UNECE Worksession on Statistical Data Confidentiality, Manchester, UK,at http://www.unece.org/stats/documents/2007/12/confidentiality/wp.19.e.pdf .
Dwork, C. and Yekhanin, S. (2008), “New Efficient Attacks on Statistical Disclosure Control Mechanisms,” Advances in Cryptology—CRYPTO 2008, to appear, also at http://research.microsoft.com/research/sv/DatabasePrivacy/dy08.pdf .
Ganta, S., Prasad, S., and Smith, A. (2008), Compositional Attacks and Auxiliary Information in Data Privacy,” ACM KDD ’08, 265-273.
Fellegi, I. P., and Holt, D. (1976), "A Systematic Approach to Automatic Edit and Imputation," *Journal of the American Statistical Association*, 71, 17-35.
Fuller, W. A. (1993), “Masking Procedures for Microdata Disclosure Limitation,” *Journal of Official Statistics*, 9**,** 383-406 (http://www.jos.nu/Articles/abstract.asp?article=92383).
Kim, J. J., and Winkler, W. E. (1995), “Masking Microdata Files,” *American Statistical Association, Proceedings of the Section on Survey Research Methods*, 114-119 (http://www.amstat.org/sections/SRMS/Proceedings/papers/1995_017.pdf , longer report http://www.census.gov/srd/papers/pdf/rr97-3.pdf) .
Little, R. J. A. and Rubin, D. B. (2002), *Statistic Analysis with Missing Data (2nd Edition)*, John Wiley: New York, N.Y.
Liu, C. (2000), “Estimation of Discrete Distributions with a Class of Simplex Constraints,” *Journal of the American Statistical Association*, 95 (449), 109-120.
Machanavajjhala, A., Kifer, D., Abowd, J. Gehrke, J., and Vilhuber, L. (2008), “Privacy: Theory meets Practice on the Map,” ICDE 2008, 277-286.
Meng, X.-L., and Rubin, D. B. (1993), “Maximum Likelihood via the ECM Algorithm: A General Framework,” *Biometrika*, 80, 267-78.
Winkler, W. E. (1990), “On Dykstra’s Iterative Fitting Procedure,” *Annals of Probability*, 18, 1410-1415.
Winkler, W. E. (1993), "Improved Decision Rules in the Fellegi-Sunter Model of Record Linkage," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 274-279 (also http://www.census.gov/srd/papers/pdf/rr93-12.pdf).
Winkler, W. E. (1998), “Re-identification Methods for Evaluating the Confidentiality of Analytically Valid Microdata,” *Research in Official Statistics*, 1, 87-104, http://www.census.gov/srd/papers/pdf/rrs2005-09.pdf .
Winkler, W. E. (2003), “A Contingency Table Model for Imputing Data Satisfying Analytic Constraints,” *American Statistical Association, Proc. Survey Research Methods Section*, CD-ROM, also http://www.census.gov/srd/papers/pdf/rrs2003-07.pdf .
Winkler, W.E. (2007), “Analytically Valid Discrete Microdata and Re-identification,” available at http://www.census.gov/srd/papers/pdf/rrs2007-19.pdf .
Winkler, W. E. (2008), “General Methods and Algorithms for Modeling and Imputing Discrete Data under a Variety of Constraints,” Statistical Research Division Report RRS2008-08, available at http://www.census.gov/srd/papers/pdf/rrs2008-08.pdf .
Xiao, X., and Tao, Y. (2008), “Output Perturbation with Query Relaxation,” VLDB, 857-869.

Appendix

Table A.1. Original Probabilities and Fitted Probabilities Indexed by Cell

Cell	Original	Fitted
0000 0 0 0 0	0.02859	0.02876
0001 0 0 0 1	0.25344	0.25328

0002	0	0	1	0	0.00172	0.00163
0003	0	0	1	1	0.00781	0.00790
0004	0	0	2	0	0.00031	0.00037
0005	0	0	2	1	0.00181	0.00175
0006	0	0	3	0	0.00042	0.00042
0007	0	0	3	1	0.00210	0.00210
0008	0	0	4	0	0.00338	0.00326
0009	0	0	4	1	0.01451	0.01463
0010	0	1	0	0	0.05369	0.05383
0011	0	1	0	1	0.03602	0.03588
0012	0	1	1	0	0.00104	0.00102
0013	0	1	1	1	0.00038	0.00040
0014	0	1	2	0	0.00062	0.00054
0015	0	1	2	1	0.00042	0.00050
0016	0	1	3	0	0.00044	0.00044
0017	0	1	3	1	0.00022	0.00022
0018	0	1	4	0	0.00714	0.00710
0019	0	1	4	1	0.00307	0.00311
0020	0	2	0	0	0.09670	0.09636
0021	0	2	0	1	0.12426	0.12460
0022	0	2	1	0	0.00369	0.00381
0023	0	2	1	1	0.00462	0.00451
0024	0	2	2	0	0.00086	0.00088
0025	0	2	2	1	0.00148	0.00146
0026	0	2	3	0	0.00104	0.00104
0027	0	2	3	1	0.00150	0.00150
0028	0	2	4	0	0.01656	0.01677
0029	0	2	4	1	0.01486	0.01466
0030	0	3	0	0	0.01030	0.01031
0031	0	3	0	1	0.00692	0.00692
0032	0	3	1	0	0.00024	0.00024
0033	0	3	1	1	0.00020	0.00020
0034	0	3	2	0	0.00022	0.00022
0035	0	3	2	1	0.00009	0.00009
0036	0	3	3	0	0.00027	0.00027
0037	0	3	3	1	0.00011	0.00011
0038	0	3	4	0	0.00418	0.00418
0039	0	3	4	1	0.00190	0.00190
0040	0	4	0	0	0.01320	0.01321
0041	0	4	0	1	0.00268	0.00267
0042	0	4	1	0	0.00046	0.00046
0043	0	4	1	1	0.00009	0.00009
0044	0	4	2	0	0.00013	0.00013
0045	0	4	2	1	0.00002	0.00000
0046	0	4	3	0	0.00009	0.00009
0047	0	4	3	1	0.00004	0.00000
0048	0	4	4	0	0.00232	0.00232
0049	0	4	4	1	0.00035	0.00036
0050	0	5	0	0	0.00318	0.00322
0051	0	5	0	1	0.00363	0.00359
0052	0	5	1	0	0.00042	0.00042
0053	0	5	1	1	0.00042	0.00042
0054	0	5	2	0	0.00007	0.00007
0055	0	5	2	1	0.00011	0.00011
0056	0	5	3	0	0.00015	0.00015
0057	0	5	3	1	0.00020	0.00020

0058	0	5	4	0	0.00088	0.00085
0059	0	5	4	1	0.00035	0.00039
0060	0	6	0	0	0.00031	0.00031
0061	0	6	0	1	0.00015	0.00015
0062	0	6	1	0	0.00002	0.00000
0063	0	6	1	1	0.00000	0.00001
0064	0	6	2	0	0.00000	0.00001
0065	0	6	2	1	0.00000	0.00001
0066	0	6	3	0	0.00000	0.00001
0067	0	6	3	1	0.00000	0.00001
0068	0	6	4	0	0.00002	0.00000
0069	0	6	4	1	0.00002	0.00000
0070	1	0	0	0	0.00394	0.00388
0071	1	0	0	1	0.04869	0.04875
0072	1	0	1	0	0.00007	0.00012
0073	1	0	1	1	0.00119	0.00114
0074	1	0	2	0	0.00002	0.00000
0075	1	0	2	1	0.00042	0.00042
0076	1	0	3	0	0.00000	0.00002
0077	1	0	3	1	0.00020	0.00020
0078	1	0	4	0	0.00015	0.00016
0079	1	0	4	1	0.00095	0.00095
0080	1	1	0	0	0.00312	0.00314
0081	1	1	0	1	0.00551	0.00548
0082	1	1	1	0	0.00009	0.00007
0083	1	1	1	1	0.00009	0.00010
0084	1	1	2	0	0.00004	0.00000
0085	1	1	2	1	0.00007	0.00007
0086	1	1	3	0	0.00000	0.00001
0087	1	1	3	1	0.00004	0.00000
0088	1	1	4	0	0.00020	0.00019
0089	1	1	4	1	0.00020	0.00021
0090	1	2	0	0	0.00281	0.00284
0091	1	2	0	1	0.00887	0.00884
0092	1	2	1	0	0.00011	0.00008
0093	1	2	1	1	0.00027	0.00030
0094	1	2	2	0	0.00004	0.00000
0095	1	2	2	1	0.00011	0.00011
0096	1	2	3	0	0.00002	0.00000
0097	1	2	3	1	0.00002	0.00000
0098	1	2	4	0	0.00031	0.00032
0099	1	2	4	1	0.00064	0.00063
0100	1	3	0	0	0.00062	0.00062
0101	1	3	0	1	0.00100	0.00100
0102	1	3	1	0	0.00000	0.00001
0103	1	3	1	1	0.00002	0.00000
0104	1	3	2	0	0.00000	0.00000
0105	1	3	2	1	0.00000	0.00001
0106	1	3	3	0	0.00000	0.00001
0107	1	3	3	1	0.00002	0.00000
0108	1	3	4	0	0.00004	0.00000
0109	1	3	4	1	0.00015	0.00015
0110	1	4	0	0	0.00153	0.00153
0111	1	4	0	1	0.00106	0.00106
0112	1	4	1	0	0.00007	0.00007
0113	1	4	1	1	0.00000	0.00006

0114	1	4	2	0	0.00000	0.00001
0115	1	4	2	1	0.00000	0.00001
0116	1	4	3	0	0.00000	0.00001
0117	1	4	3	1	0.00000	0.00001
0118	1	4	4	0	0.00009	0.00009
0119	1	4	4	1	0.00002	0.00000
0120	1	5	0	0	0.00031	0.00031
0121	1	5	0	1	0.00055	0.00055
0122	1	5	1	0	0.00000	0.00004
0123	1	5	1	1	0.00009	0.00009
0124	1	5	2	0	0.00002	0.00000
0125	1	5	2	1	0.00002	0.00000
0126	1	5	3	0	0.00000	0.00001
0127	1	5	3	1	0.00000	0.00001
0128	1	5	4	0	0.00002	0.00000
0129	1	5	4	1	0.00004	0.00000
0130	1	6	0	0	0.00002	0.00000
0131	1	6	0	1	0.00004	0.00000
0132	1	6	1	0	0.00000	0.00001
0133	1	6	1	1	0.00000	0.00001
0134	1	6	2	0	0.00000	0.00000
0135	1	6	2	1	0.00000	0.00001
0136	1	6	3	0	0.00000	0.00001
0137	1	6	3	1	0.00000	0.00001
0138	1	6	4	0	0.00000	0.00001
0139	1	6	4	1	0.00000	0.00001
0140	2	0	0	0	0.00126	0.00125
0141	2	0	0	1	0.02430	0.02431
0142	2	0	1	0	0.00015	0.00016
0143	2	0	1	1	0.00077	0.00077
0144	2	0	2	0	0.00002	0.00000
0145	2	0	2	1	0.00002	0.00000
0146	2	0	3	0	0.00000	0.00001
0147	2	0	3	1	0.00009	0.00009
0148	2	0	4	0	0.00002	0.00000
0149	2	0	4	1	0.00044	0.00044
0150	2	1	0	0	0.00093	0.00093
0151	2	1	0	1	0.00197	0.00197
0152	2	1	1	0	0.00004	0.00000
0153	2	1	1	1	0.00007	0.00007
0154	2	1	2	0	0.00000	0.00001
0155	2	1	2	1	0.00000	0.00001
0156	2	1	3	0	0.00000	0.00001
0157	2	1	3	1	0.00000	0.00001
0158	2	1	4	0	0.00002	0.00000
0159	2	1	4	1	0.00013	0.00013
0160	2	2	0	0	0.00102	0.00102
0161	2	2	0	1	0.00321	0.00320
0162	2	2	1	0	0.00007	0.00006
0163	2	2	1	1	0.00007	0.00007
0164	2	2	2	0	0.00000	0.00001
0165	2	2	2	1	0.00000	0.00001
0166	2	2	3	0	0.00000	0.00002
0167	2	2	3	1	0.00002	0.00000
0168	2	2	4	0	0.00007	0.00007
0169	2	2	4	1	0.00013	0.00013

0170	2	3	0	0	0.00018	0.00018
0171	2	3	0	1	0.00035	0.00035
0172	2	3	1	0	0.00000	0.00001
0173	2	3	1	1	0.00002	0.00000
0174	2	3	2	0	0.00000	0.00001
0175	2	3	2	1	0.00000	0.00001
0176	2	3	3	0	0.00000	0.00001
0177	2	3	3	1	0.00000	0.00001
0178	2	3	4	0	0.00000	0.00001
0179	2	3	4	1	0.00000	0.00001
0180	2	4	0	0	0.00055	0.00055
0181	2	4	0	1	0.00027	0.00027
0182	2	4	1	0	0.00000	0.00003
0183	2	4	1	1	0.00002	0.00000
0184	2	4	2	0	0.00000	0.00002
0185	2	4	2	1	0.00000	0.00000
0186	2	4	3	0	0.00000	0.00001
0187	2	4	3	1	0.00000	0.00001
0188	2	4	4	0	0.00002	0.00000
0189	2	4	4	1	0.00000	0.00001
0190	2	5	0	0	0.00004	0.00000
0191	2	5	0	1	0.00009	0.00009
0192	2	5	1	0	0.00000	0.00001
0193	2	5	1	1	0.00002	0.00000
0194	2	5	2	0	0.00000	0.00001
0195	2	5	2	1	0.00000	0.00001
0196	2	5	3	0	0.00000	0.00000
0197	2	5	3	1	0.00000	0.00002
0198	2	5	4	0	0.00000	0.00001
0199	2	5	4	1	0.00000	0.00001
0200	2	6	0	0	0.00000	0.00001
0201	2	6	0	1	0.00000	0.00001
0202	2	6	1	0	0.00000	0.00002
0203	2	6	1	1	0.00000	0.00001
0204	2	6	2	0	0.00000	0.00001
0205	2	6	2	1	0.00000	0.00001
0206	2	6	3	0	0.00000	0.00001
0207	2	6	3	1	0.00000	0.00001
0208	2	6	4	0	0.00000	0.00001
0209	2	6	4	1	0.00000	0.00001
0210	3	0	0	0	0.00093	0.00086
0211	3	0	0	1	0.01176	0.01183
0212	3	0	1	0	0.00013	0.00013
0213	3	0	1	1	0.00075	0.00075
0214	3	0	2	0	0.00007	0.00006
0215	3	0	2	1	0.00018	0.00018
0216	3	0	3	0	0.00002	0.00000
0217	3	0	3	1	0.00004	0.00000
0218	3	0	4	0	0.00018	0.00025
0219	3	0	4	1	0.00157	0.00150
0220	3	1	0	0	0.00190	0.00202
0221	3	1	0	1	0.00170	0.00159
0222	3	1	1	0	0.00009	0.00009
0223	3	1	1	1	0.00004	0.00000
0224	3	1	2	0	0.00013	0.00013
0225	3	1	2	1	0.00004	0.00000

0226	3	1	3	0	0.00002	0.00000
0227	3	1	3	1	0.00002	0.00000
0228	3	1	4	0	0.00097	0.00086
0229	3	1	4	1	0.00027	0.00038
0230	3	2	0	0	0.00243	0.00240
0231	3	2	0	1	0.00305	0.00308
0232	3	2	1	0	0.00013	0.00013
0233	3	2	1	1	0.00011	0.00011
0234	3	2	2	0	0.00011	0.00012
0235	3	2	2	1	0.00009	0.00008
0236	3	2	3	0	0.00004	0.00000
0237	3	2	3	1	0.00004	0.00000
0238	3	2	4	0	0.00108	0.00111
0239	3	2	4	1	0.00086	0.00084
0240	3	3	0	0	0.00022	0.00020
0241	3	3	0	1	0.00013	0.00015
0242	3	3	1	0	0.00000	0.00001
0243	3	3	1	1	0.00000	0.00001
0244	3	3	2	0	0.00002	0.00000
0245	3	3	2	1	0.00000	0.00000
0246	3	3	3	0	0.00002	0.00000
0247	3	3	3	1	0.00000	0.00001
0248	3	3	4	0	0.00029	0.00031
0249	3	3	4	1	0.00015	0.00014
0250	3	4	0	0	0.00064	0.00064
0251	3	4	0	1	0.00020	0.00020
0252	3	4	1	0	0.00000	0.00002
0253	3	4	1	1	0.00000	0.00000
0254	3	4	2	0	0.00009	0.00009
0255	3	4	2	1	0.00000	0.00001
0256	3	4	3	0	0.00000	0.00001
0257	3	4	3	1	0.00000	0.00001
0258	3	4	4	0	0.00011	0.00011
0259	3	4	4	1	0.00002	0.00000
0260	3	5	0	0	0.00013	0.00013
0261	3	5	0	1	0.00007	0.00007
0262	3	5	1	0	0.00000	0.00002
0263	3	5	1	1	0.00002	0.00000
0264	3	5	2	0	0.00000	0.00002
0265	3	5	2	1	0.00000	0.00001
0266	3	5	3	0	0.00000	0.00001
0267	3	5	3	1	0.00000	0.00001
0268	3	5	4	0	0.00002	0.00000
0269	3	5	4	1	0.00009	0.00009
0270	3	6	0	0	0.00004	0.00000
0271	3	6	0	1	0.00002	0.00000
0272	3	6	1	0	0.00000	0.00001
0273	3	6	1	1	0.00000	0.00001
0274	3	6	2	0	0.00000	0.00001
0275	3	6	2	1	0.00000	0.00001
0276	3	6	3	0	0.00000	0.00001
0277	3	6	3	1	0.00000	0.00001
0278	3	6	4	0	0.00000	0.00001
0279	3	6	4	1	0.00000	0.00001
0280	4	0	0	0	0.00433	0.00428
0281	4	0	0	1	0.02508	0.02513

0282	4	0	1	0	0.00004	0.00000
0283	4	0	1	1	0.00055	0.00055
0284	4	0	2	0	0.00015	0.00010
0285	4	0	2	1	0.00029	0.00034
0286	4	0	3	0	0.00000	0.00002
0287	4	0	3	1	0.00013	0.00013
0288	4	0	4	0	0.00053	0.00064
0289	4	0	4	1	0.00248	0.00237
0290	4	1	0	0	0.00705	0.00685
0291	4	1	0	1	0.00230	0.00250
0292	4	1	1	0	0.00009	0.00013
0293	4	1	1	1	0.00009	0.00005
0294	4	1	2	0	0.00015	0.00023
0295	4	1	2	1	0.00020	0.00012
0296	4	1	3	0	0.00002	0.00000
0297	4	1	3	1	0.00000	0.00001
0298	4	1	4	0	0.00119	0.00128
0299	4	1	4	1	0.00046	0.00038
0300	4	2	0	0	0.00688	0.00715
0301	4	2	0	1	0.00637	0.00610
0302	4	2	1	0	0.00020	0.00016
0303	4	2	1	1	0.00020	0.00024
0304	4	2	2	0	0.00029	0.00026
0305	4	2	2	1	0.00027	0.00029
0306	4	2	3	0	0.00002	0.00000
0307	4	2	3	1	0.00004	0.00000
0308	4	2	4	0	0.00208	0.00187
0309	4	2	4	1	0.00113	0.00133
0310	4	3	0	0	0.00069	0.00071
0311	4	3	0	1	0.00049	0.00046
0312	4	3	1	0	0.00000	0.00001
0313	4	3	1	1	0.00000	0.00001
0314	4	3	2	0	0.00002	0.00000
0315	4	3	2	1	0.00000	0.00001
0316	4	3	3	0	0.00002	0.00000
0317	4	3	3	1	0.00002	0.00000
0318	4	3	4	0	0.00060	0.00057
0319	4	3	4	1	0.00029	0.00031
0320	4	4	0	0	0.00215	0.00214
0321	4	4	0	1	0.00035	0.00036
0322	4	4	1	0	0.00007	0.00007
0323	4	4	1	1	0.00002	0.00000
0324	4	4	2	0	0.00007	0.00007
0325	4	4	2	1	0.00000	0.00001
0326	4	4	3	0	0.00000	0.00001
0327	4	4	3	1	0.00000	0.00000
0328	4	4	4	0	0.00038	0.00038
0329	4	4	4	1	0.00007	0.00006
0330	4	5	0	0	0.00031	0.00028
0331	4	5	0	1	0.00013	0.00017
0332	4	5	1	0	0.00002	0.00000
0333	4	5	1	1	0.00000	0.00001
0334	4	5	2	0	0.00000	0.00001
0335	4	5	2	1	0.00000	0.00001
0336	4	5	3	0	0.00000	0.00001
0337	4	5	3	1	0.00000	0.00001

0338	4	5	4	0	0.00015	0.00019
0339	4	5	4	1	0.00009	0.00006
0340	4	6	0	0	0.00000	0.00001
0341	4	6	0	1	0.00000	0.00001
0342	4	6	1	0	0.00000	0.00001
0343	4	6	1	1	0.00000	0.00001
0344	4	6	2	0	0.00000	0.00001
0345	4	6	2	1	0.00000	0.00001
0346	4	6	3	0	0.00000	0.00001
0347	4	6	3	1	0.00000	0.00001
0348	4	6	4	0	0.00000	0.00001
0349	4	6	4	1	0.00000	0.00001
0350	5	0	0	0	0.00179	0.00181
0351	5	0	0	1	0.01468	0.01466
0352	5	0	1	0	0.00011	0.00014
0353	5	0	1	1	0.00088	0.00085
0354	5	0	2	0	0.00004	0.00000
0355	5	0	2	1	0.00022	0.00022
0356	5	0	3	0	0.00002	0.00000
0357	5	0	3	1	0.00007	0.00007
0358	5	0	4	0	0.00042	0.00037
0359	5	0	4	1	0.00102	0.00107
0360	5	1	0	0	0.00374	0.00366
0361	5	1	0	1	0.00190	0.00198
0362	5	1	1	0	0.00015	0.00015
0363	5	1	1	1	0.00002	0.00000
0364	5	1	2	0	0.00007	0.00007
0365	5	1	2	1	0.00002	0.00000
0366	5	1	3	0	0.00007	0.00007
0367	5	1	3	1	0.00002	0.00000
0368	5	1	4	0	0.00066	0.00074
0369	5	1	4	1	0.00027	0.00019
0370	5	2	0	0	0.00551	0.00557
0371	5	2	0	1	0.00540	0.00533
0372	5	2	1	0	0.00029	0.00025
0373	5	2	1	1	0.00027	0.00030
0374	5	2	2	0	0.00002	0.00000
0375	5	2	2	1	0.00007	0.00007
0376	5	2	3	0	0.00007	0.00007
0377	5	2	3	1	0.00004	0.00000
0378	5	2	4	0	0.00155	0.00152
0379	5	2	4	1	0.00066	0.00069
0380	5	3	0	0	0.00062	0.00061
0381	5	3	0	1	0.00038	0.00038
0382	5	3	1	0	0.00002	0.00000
0383	5	3	1	1	0.00002	0.00000
0384	5	3	2	0	0.00000	0.00001
0385	5	3	2	1	0.00000	0.00001
0386	5	3	3	0	0.00000	0.00001
0387	5	3	3	1	0.00000	0.00000
0388	5	3	4	0	0.00027	0.00027
0389	5	3	4	1	0.00009	0.00008
0390	5	4	0	0	0.00073	0.00073
0391	5	4	0	1	0.00004	0.00000
0392	5	4	1	0	0.00002	0.00000
0393	5	4	1	1	0.00000	0.00000

0394	5	4	2	0	0.00004	0.00000
0395	5	4	2	1	0.00000	0.00000
0396	5	4	3	0	0.00002	0.00000
0397	5	4	3	1	0.00000	0.00000
0398	5	4	4	0	0.00013	0.00013
0399	5	4	4	1	0.00004	0.00000
0400	5	5	0	0	0.00018	0.00018
0401	5	5	0	1	0.00018	0.00018
0402	5	5	1	0	0.00007	0.00007
0403	5	5	1	1	0.00002	0.00000
0404	5	5	2	0	0.00002	0.00000
0405	5	5	2	1	0.00000	0.00001
0406	5	5	3	0	0.00000	0.00001
0407	5	5	3	1	0.00000	0.00001
0408	5	5	4	0	0.00002	0.00000
0409	5	5	4	1	0.00002	0.00000
0410	5	6	0	0	0.00004	0.00000
0411	5	6	0	1	0.00000	0.00001
0412	5	6	1	0	0.00000	0.00001
0413	5	6	1	1	0.00000	0.00001
0414	5	6	2	0	0.00000	0.00001
0415	5	6	2	1	0.00000	0.00001
0416	5	6	3	0	0.00000	0.00001
0417	5	6	3	1	0.00000	0.00001
0418	5	6	4	0	0.00000	0.00001
0419	5	6	4	1	0.00000	0.00001
0420	6	0	0	0	0.00009	0.00009
0421	6	0	0	1	0.00018	0.00018
0422	6	0	1	0	0.00000	0.00001
0423	6	0	1	1	0.00002	0.00000
0424	6	0	2	0	0.00000	0.00001
0425	6	0	2	1	0.00000	0.00001
0426	6	0	3	0	0.00000	0.00001
0427	6	0	3	1	0.00000	0.00002
0428	6	0	4	0	0.00000	0.00001
0429	6	0	4	1	0.00000	0.00001
0430	6	1	0	0	0.00000	0.00001
0431	6	1	0	1	0.00000	0.00001
0432	6	1	1	0	0.00000	0.00001
0433	6	1	1	1	0.00000	0.00001
0434	6	1	2	0	0.00000	0.00001
0435	6	1	2	1	0.00000	0.00001
0436	6	1	3	0	0.00000	0.00001
0437	6	1	3	1	0.00000	0.00001
0438	6	1	4	0	0.00000	0.00001
0439	6	1	4	1	0.00000	0.00001
0440	6	2	0	0	0.00002	0.00000
0441	6	2	0	1	0.00009	0.00009
0442	6	2	1	0	0.00000	0.00000
0443	6	2	1	1	0.00000	0.00001
0444	6	2	2	0	0.00000	0.00000
0445	6	2	2	1	0.00000	0.00001
0446	6	2	3	0	0.00000	0.00002
0447	6	2	3	1	0.00000	0.00013
0448	6	2	4	0	0.00000	0.00001
0449	6	2	4	1	0.00002	0.00000

0450	6	3	0	0	0.00000	0.00001
0451	6	3	0	1	0.00000	0.00001
0452	6	3	1	0	0.00000	0.00001
0453	6	3	1	1	0.00000	0.00001
0454	6	3	2	0	0.00000	0.00001
0455	6	3	2	1	0.00000	0.00001
0456	6	3	3	0	0.00000	0.00001
0457	6	3	3	1	0.00000	0.00001
0458	6	3	4	0	0.00000	0.00001
0459	6	3	4	1	0.00000	0.00001
0460	6	4	0	0	0.00002	0.00000
0461	6	4	0	1	0.00000	0.00001
0462	6	4	1	0	0.00000	0.00001
0463	6	4	1	1	0.00000	0.00001
0464	6	4	2	0	0.00000	0.00001
0465	6	4	2	1	0.00000	0.00001
0466	6	4	3	0	0.00000	0.00001
0467	6	4	3	1	0.00000	0.00002
0468	6	4	4	0	0.00000	0.00001
0469	6	4	4	1	0.00000	0.00001
0470	6	5	0	0	0.00002	0.00000
0471	6	5	0	1	0.00000	0.00001
0472	6	5	1	0	0.00000	0.00001
0473	6	5	1	1	0.00000	0.00001
0474	6	5	2	0	0.00000	0.00001
0475	6	5	2	1	0.00000	0.00001
0476	6	5	3	0	0.00000	0.00001
0477	6	5	3	1	0.00000	0.00002
0478	6	5	4	0	0.00000	0.00001
0479	6	5	4	1	0.00000	0.00001
0480	6	6	0	0	0.00000	0.00001
0481	6	6	0	1	0.00000	0.00001
0482	6	6	1	0	0.00000	0.00001
0483	6	6	1	1	0.00000	0.00001
0484	6	6	2	0	0.00000	0.00001
0485	6	6	2	1	0.00000	0.00001
0486	6	6	3	0	0.00000	0.00001
0487	6	6	3	1	0.00000	0.00001
0488	6	6	4	0	0.00000	0.00001
0489	6	6	4	1	0.00000	0.00001

Maxdiff1= 0.000340 at pattern=0020
Maxdiff2= 0.000340 at pattern=0021
Maxdiff3= 0.000272 at pattern=0301

Table A.2. 3-way Margins associated with Original and Fitted Data

Pattern = 0, Variables 1,2,3

00000	0.282037	0.282037
00001	0.009531	0.009531
00002	0.002123	0.002123
00003	0.002521	0.002521
00004	0.017890	0.017890

00005	0.089715	0.089715
00006	0.001415	0.001415
00007	0.001039	0.001039
00008	0.000663	0.000663
00009	0.010216	0.010216
00010	0.220959	0.220959
00011	0.008315	0.008315
00012	0.002344	0.002344
00013	0.002543	0.002544
00014	0.031423	0.031423
00015	0.017227	0.017227
00016	0.000442	0.000442
00017	0.000310	0.000310
00018	0.000376	0.000376
00019	0.006081	0.006081
00020	0.015878	0.015878
00021	0.000553	0.000553
00022	0.000155	0.000137
00023	0.000133	0.000092
00024	0.002676	0.002676
00025	0.006811	0.006811
00026	0.000840	0.000840
00027	0.000177	0.000177
00028	0.000354	0.000354
00029	0.001238	0.001238
00030	0.000464	0.000464
00031	0.000022	0.000010
00032	0.000000	0.000018
00033	0.000000	0.000019
00034	0.000044	0.000007
00035	0.052630	0.052630
00036	0.001260	0.001260
00037	0.000442	0.000424
00038	0.000199	0.000219
00039	0.001106	0.001106
00040	0.008624	0.008624
00041	0.000177	0.000177
00042	0.000111	0.000070
00043	0.000044	0.000010
00044	0.000398	0.000398
00045	0.011676	0.011676
00046	0.000376	0.000376
00047	0.000155	0.000114
00048	0.000044	0.000008
00049	0.000951	0.000951
00050	0.001614	0.001614
00051	0.000022	0.000009
00052	0.000000	0.000014
00053	0.000022	0.000011
00054	0.000199	0.000159
00055	0.002587	0.002587
00056	0.000066	0.000127
00057	0.000000	0.000013
00058	0.000000	0.000019
00059	0.000111	0.000092
00060	0.000862	0.000862

00061	0.000088	0.000128
00062	0.000044	0.000006
00063	0.000000	0.000019
00064	0.000066	0.000007
00065	0.000066	0.000007
00066	0.000000	0.000018
00067	0.000000	0.000015
00068	0.000000	0.000018
00069	0.000000	0.000020
00070	0.025563	0.025563
00071	0.000929	0.000929
00072	0.000044	0.000006
00073	0.000088	0.000096
00074	0.000464	0.000446
00075	0.002897	0.002897
00076	0.000111	0.000070
00077	0.000000	0.000019
00078	0.000000	0.000017
00079	0.000155	0.000137
00080	0.004224	0.004224
00081	0.000133	0.000133
00082	0.000000	0.000018
00083	0.000022	0.000026
00084	0.000199	0.000199
00085	0.000531	0.000531
00086	0.000022	0.000015
00087	0.000000	0.000018
00088	0.000000	0.000016
00089	0.000000	0.000020
00090	0.000818	0.000818
00091	0.000022	0.000030
00092	0.000000	0.000023
00093	0.000000	0.000018
00094	0.000022	0.000009
00095	0.000133	0.000092
00096	0.000022	0.000013
00097	0.000000	0.000018
00098	0.000000	0.000019
00099	0.000000	0.000018
00100	0.000000	0.000018
00101	0.000000	0.000027
00102	0.000000	0.000020
00103	0.000000	0.000018
00104	0.000000	0.000021
00105	0.012693	0.012693
00106	0.000885	0.000885
00107	0.000243	0.000243
00108	0.000066	0.000006
00109	0.001747	0.001747
00110	0.003605	0.003605
00111	0.000133	0.000092
00112	0.000177	0.000137
00113	0.000044	0.000007
00114	0.001238	0.001238
00115	0.005484	0.005484
00116	0.000243	0.000243

00117	0.000199	0.000199
00118	0.000088	0.000008
00119	0.001946	0.001946
00120	0.000354	0.000354
00121	0.000000	0.000020
00122	0.000022	0.000007
00123	0.000022	0.000010
00124	0.000442	0.000442
00125	0.000840	0.000840
00126	0.000000	0.000021
00127	0.000088	0.000098
00128	0.000000	0.000017
00129	0.000133	0.000114
00130	0.000199	0.000199
00131	0.000022	0.000021
00132	0.000000	0.000022
00133	0.000000	0.000019
00134	0.000111	0.000092
00135	0.000066	0.000007
00136	0.000000	0.000020
00137	0.000000	0.000020
00138	0.000000	0.000019
00139	0.000000	0.000019
00140	0.029411	0.029411
00141	0.000597	0.000557
00142	0.000442	0.000442
00143	0.000133	0.000149
00144	0.003007	0.003007
00145	0.009354	0.009354
00146	0.000177	0.000177
00147	0.000354	0.000354
00148	0.000022	0.000009
00149	0.001659	0.001659
00150	0.013246	0.013246
00151	0.000398	0.000398
00152	0.000553	0.000553
00153	0.000066	0.000008
00154	0.003206	0.003206
00155	0.001172	0.001172
00156	0.000000	0.000017
00157	0.000022	0.000009
00158	0.000044	0.000007
00159	0.000885	0.000885
00160	0.002499	0.002499
00161	0.000088	0.000070
00162	0.000066	0.000076
00163	0.000000	0.000013
00164	0.000442	0.000442
00165	0.000442	0.000442
00166	0.000022	0.000012
00167	0.000000	0.000017
00168	0.000000	0.000015
00169	0.000243	0.000243
00170	0.000000	0.000019
00171	0.000000	0.000018
00172	0.000000	0.000019

00173	0.000000	0.000017
00174	0.000000	0.000018
00175	0.016475	0.016475
00176	0.000995	0.000995
00177	0.000265	0.000225
00178	0.000088	0.000068
00179	0.001437	0.001437
00180	0.005639	0.005639
00181	0.000177	0.000159
00182	0.000088	0.000070
00183	0.000088	0.000070
00184	0.000929	0.000929
00185	0.010902	0.010902
00186	0.000553	0.000553
00187	0.000088	0.000070
00188	0.000111	0.000070
00189	0.002211	0.002211
00190	0.000995	0.000995
00191	0.000044	0.000007
00192	0.000000	0.000017
00193	0.000000	0.000013
00194	0.000354	0.000354
00195	0.000774	0.000734
00196	0.000022	0.000005
00197	0.000044	0.000005
00198	0.000022	0.000003
00199	0.000177	0.000136
00200	0.000354	0.000354
00201	0.000088	0.000070
00202	0.000022	0.000016
00203	0.000000	0.000014
00204	0.000044	0.000007
00205	0.000044	0.000013
00206	0.000000	0.000018
00207	0.000000	0.000018
00208	0.000000	0.000014
00209	0.000000	0.000018
00210	0.000265	0.000265
00211	0.000022	0.000015
00212	0.000000	0.000022
00213	0.000000	0.000024
00214	0.000000	0.000023
00215	0.000000	0.000019
00216	0.000000	0.000018
00217	0.000000	0.000019
00218	0.000000	0.000020
00219	0.000000	0.000018
00220	0.000111	0.000092
00221	0.000000	0.000018
00222	0.000000	0.000018
00223	0.000000	0.000145
00224	0.000022	0.000009
00225	0.000000	0.000019
00226	0.000000	0.000019
00227	0.000000	0.000018
00228	0.000000	0.000019

00229 0.000000 0.000018
00230 0.000022 0.000014
00231 0.000000 0.000018
00232 0.000000 0.000018
00233 0.000000 0.000025
00234 0.000000 0.000017
00235 0.000022 0.000014
00236 0.000000 0.000018
00237 0.000000 0.000018
00238 0.000000 0.000022
00239 0.000000 0.000019
00240 0.000000 0.000018
00241 0.000000 0.000018
00242 0.000000 0.000018
00243 0.000000 0.000021
00244 0.000000 0.000019
Maxdiff1= 0.000145 at pattern=0223
Maxdiff2= 0.000080 at pattern=0118
Maxdiff3= 0.000061 at pattern=0056

Pattern = 1, Variables 1,2,4
00000 0.034431 0.034431
00001 0.279671 0.279671
00002 0.062935 0.062935
00003 0.040114 0.040114
00004 0.118861 0.118861
00005 0.146724 0.146724
00006 0.015214 0.015214
00007 0.009221 0.009221
00008 0.016209 0.016209
00009 0.003184 0.003126
00010 0.004710 0.004710
00011 0.004710 0.004710
00012 0.000354 0.000341
00013 0.000177 0.000179
00014 0.004179 0.004182
00015 0.051458 0.051458
00016 0.003450 0.003415
00017 0.005904 0.005864
00018 0.003295 0.003236
00019 0.009907 0.009889
00020 0.000663 0.000640
00021 0.001194 0.001167
00022 0.001681 0.001695
00023 0.001084 0.001145
00024 0.000354 0.000360
00025 0.000708 0.000663
00026 0.000022 0.000033
00027 0.000044 0.000045
00028 0.001459 0.001429
00029 0.025630 0.025611
00030 0.000995 0.000951
00031 0.002167 0.002189
00032 0.001150 0.001178
00033 0.003428 0.003421
00034 0.000177 0.000216

00035	0.000376	0.000385
00036	0.000575	0.000613
00037	0.000287	0.000286
00038	0.000044	0.000030
00039	0.000111	0.000130
00040	0.000000	0.000059
00041	0.000000	0.000045
00042	0.001327	0.001306
00043	0.014308	0.014267
00044	0.003118	0.003099
00045	0.002079	0.001980
00046	0.003804	0.003763
00047	0.004157	0.004117
00048	0.000553	0.000527
00049	0.000287	0.000306
00050	0.000840	0.000868
00051	0.000221	0.000223
00052	0.000155	0.000181
00053	0.000177	0.000172
00054	0.000044	0.000044
00055	0.000022	0.000041
00056	0.005064	0.005040
00057	0.028527	0.028527
00058	0.008514	0.008495
00059	0.003052	0.003057
00060	0.009465	0.009447
00061	0.008005	0.007965
00062	0.001327	0.001297
00063	0.000796	0.000794
00064	0.002654	0.002662
00065	0.000442	0.000438
00066	0.000486	0.000483
00067	0.000221	0.000247
00068	0.000000	0.000038
00069	0.000000	0.000053
00070	0.002388	0.002328
00071	0.016873	0.016873
00072	0.004688	0.004688
00073	0.002233	0.002179
00074	0.007430	0.007412
00075	0.006435	0.006395
00076	0.000907	0.000907
00077	0.000486	0.000479
00078	0.000951	0.000872
00079	0.000088	0.000011
00080	0.000287	0.000256
00081	0.000221	0.000206
00082	0.000044	0.000034
00083	0.000000	0.000048
00084	0.000088	0.000132
00085	0.000199	0.000216
00086	0.000000	0.000047
00087	0.000000	0.000047
00088	0.000022	0.000038
00089	0.000111	0.000245
00090	0.000000	0.000047

00091 0.000000 0.000046
00092 0.000022 0.000040
00093 0.000000 0.000052
00094 0.000022 0.000039
00095 0.000000 0.000051
00096 0.000000 0.000044
00097 0.000000 0.000051
Maxdiff1= 0.000134 at pattern=0089
Maxdiff2= 0.000099 at pattern=0045
Maxdiff3= 0.000079 at pattern=0078

Pattern = 2, Variables 2,3,4
00000 0.040932 0.040932
00001 0.378143 0.378143
00002 0.002233 0.002205
00003 0.011986 0.011967
00004 0.000619 0.000554
00005 0.002941 0.002932
00006 0.000464 0.000477
00007 0.002632 0.002606
00008 0.004688 0.004681
00009 0.020964 0.020975
00010 0.070432 0.070440
00011 0.049402 0.049412
00012 0.001504 0.001475
00013 0.000686 0.000633
00014 0.001017 0.000993
00015 0.000752 0.000715
00016 0.000553 0.000536
00017 0.000310 0.000261
00018 0.010194 0.010186
00019 0.004401 0.004409
00020 0.115367 0.115348
00021 0.151235 0.151235
00022 0.004489 0.004494
00023 0.005528 0.005541
00024 0.001327 0.001279
00025 0.002012 0.002038
00026 0.001194 0.001160
00027 0.001681 0.001650
00028 0.021649 0.021654
00029 0.018310 0.018292
00030 0.012627 0.012635
00031 0.009266 0.009276
00032 0.000265 0.000291
00033 0.000265 0.000239
00034 0.000265 0.000267
00035 0.000088 0.000126
00036 0.000310 0.000303
00037 0.000155 0.000151
00038 0.005374 0.005352
00039 0.002587 0.002607
00040 0.018819 0.018800
00041 0.004600 0.004569
00042 0.000619 0.000654
00043 0.000133 0.000171

00044	0.000332	0.000326
00045	0.000022	0.000043
00046	0.000111	0.000135
00047	0.000044	0.000052
00048	0.003052	0.003042
00049	0.000509	0.000445
00050	0.004179	0.004121
00051	0.004644	0.004654
00052	0.000509	0.000567
00053	0.000575	0.000536
00054	0.000111	0.000111
00055	0.000133	0.000163
00056	0.000155	0.000191
00057	0.000199	0.000271
00058	0.001106	0.001071
00059	0.000597	0.000554
00060	0.000420	0.000349
00061	0.000221	0.000198
00062	0.000022	0.000063
00063	0.000000	0.000067
00064	0.000000	0.000067
00065	0.000000	0.000062
00066	0.000000	0.000054
00067	0.000000	0.000072
00068	0.000022	0.000060
00069	0.000022	0.000063

Maxdiff1= 0.000072 at pattern=0067
Maxdiff2= 0.000072 at pattern=0057
Maxdiff3= 0.000071 at pattern=0060

Pattern = 3, Variables 1,3,4

00000	0.205988	0.205988
00001	0.427102	0.427102
00002	0.007607	0.007589
00003	0.013511	0.013518
00004	0.002211	0.002223
00005	0.003936	0.003925
00006	0.002410	0.002423
00007	0.004179	0.004146
00008	0.034497	0.034479
00009	0.035072	0.035054
00010	0.012339	0.012321
00011	0.065722	0.065681
00012	0.000332	0.000384
00013	0.001659	0.001712
00014	0.000133	0.000028
00015	0.000619	0.000629
00016	0.000022	0.000059
00017	0.000287	0.000246
00018	0.000818	0.000770
00019	0.002012	0.001964
00020	0.003980	0.003950
00021	0.030185	0.030194
00022	0.000265	0.000290
00023	0.000973	0.000926
00024	0.000022	0.000065

00025	0.000022	0.000057
00026	0.000000	0.000065
00027	0.000111	0.000145
00028	0.000133	0.000105
00029	0.000708	0.000745
00030	0.006302	0.006262
00031	0.016939	0.016921
00032	0.000354	0.000412
00033	0.000929	0.000890
00034	0.000420	0.000432
00035	0.000310	0.000295
00036	0.000111	0.000038
00037	0.000111	0.000046
00038	0.002654	0.002645
00039	0.002963	0.002954
00040	0.021406	0.021416
00041	0.034718	0.034727
00042	0.000420	0.000375
00043	0.000862	0.000875
00044	0.000686	0.000685
00045	0.000752	0.000785
00046	0.000066	0.000047
00047	0.000199	0.000172
00048	0.004931	0.004939
00049	0.004511	0.004522
00050	0.012605	0.012564
00051	0.022578	0.022547
00052	0.000663	0.000634
00053	0.001216	0.001174
00054	0.000199	0.000099
00055	0.000310	0.000322
00056	0.000177	0.000155
00057	0.000133	0.000097
00058	0.003052	0.003043
00059	0.002101	0.002050
00060	0.000155	0.000125
00061	0.000265	0.000316
00062	0.000000	0.000065
00063	0.000022	0.000059
00064	0.000000	0.000065
00065	0.000000	0.000065
00066	0.000000	0.000065
00067	0.000000	0.000210
00068	0.000000	0.000065
00069	0.000022	0.000057

Maxdiff1= 0.000210 at pattern=0067
Maxdiff2= 0.000105 at pattern=0014
Maxdiff3= 0.000100 at pattern=0054