RESEARCH REPORT SERIES
(*Statistics #2003-04)*

Error Localization and Implied Edit Generation for Ratio and Balancing Edits

Maria Garcia

Statistical Research Division
U.S. Bureau of the Census
Washington D.C. 20233

Report Issued: September 2, 2003

Disclaimer: This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

Error Localization and Implied Edit Generation for Ratio and Balancing Edits

Maria Garcia
U.S. Bureau of the Census, Washington, D.C., 20233
Maria.M.Garcia@census.gov

Abstract

The U.S. Census Bureau has developed SPEER software that applies the Fellegi-Holt editing method to economic establishment surveys under ratio edit and a limited form of balancing. It is known that more than 99% of economic data only require these basic forms of edits. If implicit edits are available, then Fellegi-Holt methods have the advantage that they determine the minimal number of fields to change (error localize) so that a record satisfies all edits in one pass through the data. In most situations, implicit edits are not generated because the generation requires days-to-months of computation. In some situations when implicit edits are not available Fellegi-Holt systems use pure integer programming methods to solve the error localization problem directly and slowly (1-100 seconds per record). With only a small subset of the needed implicit edits, the current version of SPEER (Draper and Winkler 1997, upwards of 1000 records per second) applies ad hoc heuristics that finds error-localization solutions that are not optimal for as much as five percent of the edit-failing records. To maintain the speed of SPEER and do a better job of error localization, we apply the Fourier-Motzkin method to generate a large subset of the implied edits prior to error localization. In this paper, we describe the theory, computational algorithms, and results from evaluating the feasibility of this approach.

Keywords: editing, error localization, Fellegi-Holt model

1. Introduction

In economic surveys and censuses, survey data files may contain a large number of records with erroneous, missing, or inconsistent data. Errors can arise during data collection due to item non-response, misunderstanding of a survey question or problems with computer data entry. Records with erroneous or inconsistent data must be edited before the agency produces and publishes relevant and accurate statistics. Data editing is the process of identifying and correcting errors or inconsistencies in the collected survey data. In statistical agencies, data editing uses a considerable amount of the survey resources available for the publication of statistics. This cost can be reduced if we have an automated system that can be reused by various separate surveys. Currently, for most surveys, the detection and correction of erroneous data is done using an automated software. Fellegi and Holt (Fellegi and Holt, 1976) provided the theory and methodology for the creation of such a system.

An automated system based on the Fellegi-Holt methodology must satisfy the following three requirements (Fellegi and Holt, 1976):

1. The data in each record should be made to satisfy the edits by changing the fewest possible

fields.
2. The imputation rules should derive automatically from the edit rules.
3. Imputation should maintain the joint distribution of the variables (fields).

This model requires that the data in each record should be made to satisfy all edits by identifying and changing the minimum possible fields (number one above.) This criterion is referred to as the error localization problem. Fellegi and Holt showed that the implicit edits that can be logically derived from the set of analyst's supplied explicit edits are needed for solving the error localization problem. The complete set of explicit and implicit edits is sufficient to determine imputation intervals for erroneous fields so that an edit failing record is corrected. Prior edit models would fail because they lack the needed information about the original set of explicit edits that may not fail but might fail the imputed record if information in the complete set of edits is not used during error localization.

Several Fellegi-Holt computer systems are currently available for editing continuous economic data: Statistics Canada's Generalized Edit and Imputation System (GEIS) (Schiopu-Kratina and Kovar, 1989), Statistics Netherlands CherryPi (De Waal, 1996), National Agricultural Statistics Service's AGGIES (Todaro, 1999) and the US Census Bureau's Structured Program for Economic Editing and Referrals (SPEER, Draper and Winkler (1997)). The GEIS, CherryPi and AGGIES software solve simultaneous linear inequality edits using a modified Chernikova's algorithm (Rubin, 1975) to implicitly generate the failing implied edits needed for finding error localization solutions. The SPEER system is used for economic data under balancing and ratio edits and applies simple heuristics to generate a subset of the implicit edits needed for solving the error localization problem. A more detailed description of the SPEER software is given in the next section.

In this paper we applied the Fourier-Motzkin elimination method (Duffin, 1974) to generate a large subset of the implicit edits prior to error localization in the SPEER editing system. In the following sections we present the theory, computational algorithms, and results from using this approach.

2. Implicit Edit Generation and the SPEER edit system

2.1 The SPEER editing software

The Census Bureau has an editing system, SPEER (Structured Programs for Economic Editing and Referrals), for editing continuous economic data that must satisfy ratio edits and a limited form of balancing. The SPEER system has been used at the Census Bureau on several economic surveys since the early 1980's (Greenberg and Surdi1984; Greenberg and Petkunas, 1990).

This paper describes modifications to the SPEER edit software that maintain the exceptional speed of the system and do a better job of error localization. The current version of SPEER consists of a main edit program and four auxiliary modules. The FORTRAN code for the edit checking, error localization, and imputation routines in the main edit program is new. The four auxiliary modules perform different tasks: the first module automatically determines the bounds for the ratio edits (Thompson and Sigman, 1996); the second module checks the logical consistency of the user supplied explicit edits and generates the implicit ratio edits needed for error localization; the third module generates the regression coefficients that are used in the imputation module; and a new fourth module generates a subset of the implicit linear inequality edits that arise when combining ratio edits and balance equations.

The SPEER software identifies and corrects erroneous fields in data records that must satisfy ratio edits and single level balancing. By single level balancing we mean that data fields (details and totals) are allowed to be restricted by at most one balance equation. It is known that only ratio and balancing edits are required in more than 99% of economic surveys.

A record with n data fields in a computer data file is represented by $v = (v_1, v_2, \ldots, v_n)$. A ratio edit is the requirement that the ratio of two data items is bounded by lower and upper bounds,

$$l_{ij} \leq v_i / v_j \leq u_{ij},$$

where l_{ij} and u_{ij} are the largest lower bound and smallest upper bound respectively. The bounds can be determined by analysts through use of prior survey data. A balance edit is the requirement that two or more details and a reported total satisfy an additivity condition of the form

$$\sum_{k \in S} v_k - v_t = 0,$$

where S is a proper subset of the first n integers and $t \notin S$. The $v_k, k \in S$ are known as details and v_t is known as the total.

Fellegi-Holt editing model guarantees that, if the complete set of explicit and implicit edits is available then we can determine a minimum number of fields to change so that an edit failing record satisfies the edits. In the earliest versions of SPEER which used ratio edits only, it is straightforward to generate the complete set of ratio edits. Since the complete set of explicit and implicit edits is available, it is easy and exceptionally fast to solve the error localization problem.

In the most recent version of SPEER (SPEER'97), Draper and Winkler (1997) generate implicit edits induced by failing ratio edits and balance equations "on the fly" for every edit failing record. The induced edits are then used to further restrict imputation intervals than the restrictions placed by ratio edits only. The solution however, is not necessarily an error localization solution since not all implicit edits are available. This is true in most cases: in general for continuous data it is not possible to a priori generate all the implicit edits for a set of explicit linear inequality edits due to the exponential growth of the total number of implicit linear inequality edits (Sande, 1978). Recently, Winkler and Chen (2002) provided extensions to the theory and computational aspects of the Fellegi-Holt editing model for discrete data. In their research on discrete data they showed that if most of the implicit edits are computed prior to automatic editing, then error localization algorithms are faster than direct integer programming methods for solving the error localization problem. These results can be extended to continuous data. The main purpose of this paper is to use this idea in SPEER editing when a large subset, but not all, of the implicit edits are generated prior to editing.

2.2 Implicit Edit Generation for Balancing and Ratio Edits

The SPEER edit system has an auxiliary module for generating all the implicit ratio edits for a given set of explicit ratio edits. In the earlier version of SPEER (SPEER'97), the needed implicit edits implied by failing ratio edits and a balance equation are generated on the main program for every failing record. This means many implicit edits are repeatedly computed. The new SPEER software (SPEER'02) generates a large subset of the implied edits prior to SPEER editing. The implied edits are then available to be used in the main edit program. It is not necessary to repeatedly generate the same implicit edits as additional edit failing records are encountered. This eliminates the need for implicit edit generation during the computationally intensive error localization program. We want to point out that in most situations implicit edits are not generated because the generation requires days-to-months of computation, however it is feasible to generate implicit edits for SPEER algorithms because it deals with numeric data under ratio edits and single level balancing only.

The new added module for generating implicit linear inequality edits for ratio edits and balancing edits is based on the Fourier-Motzkin elimination method (Duffin, 1974). This methodology has been used in new algorithms for the Leo editing system developed at Statistics Netherlands (Quere, 2000). The Leo software uses Fourier-Motzkin elimination to delete a field from nodes representing the current set of edits in a tree search algorithm for solving the error localization problem.

The mathematical knowledge to develop and understand the implicit edit generation is simple. The method developed by Fourier for checking the consistency of a set of inequalities can be used to generate implicit linear inequality edits. Suppose we have a ratio edit,

$$l_{ij} \le v_i / v_j \le u_{ij},$$

and balance equation,

$$\sum_{k \in S} v_k - v_t = 0.$$

Using simple algebra we can rewrite the ratio edit as two linear inequality edits and the balance equation as two linear inequality edits. If we can find a variable in common in the linear inequality edits corresponding to the ratio and balance edits, say $k = i$ for some $k \in S,$ and provided the coefficients of the common variable have opposite signs, then we can eliminate the common variable by creating a linear combination of the two edits. For example, if $-v_1 + l_{14} v_4 \le 0$ and $v_1 + v_2 - v_3 \le 0$ are linear inequality edits derived from the ratio and balance equation respectively, then $v_2 - v_3 + l_{14} v_4 \le 0$ is a new implied edit. The new SPEER implicit edit generation algorithm uses this methodology to generate as many implicit edits as possible from linear combinations of the complete set of ratio edits and the balance equations. The algorithm is repeated to generate new implied edits from linear combinations of the newly generated implicit edits and the set of ratio edits. Generating a large subset of the implicit edits using this methodology has numerous advantages. For ratio edits and single level balancing the edit generation logic is simple. If implicit edits are available the speed of the main edit program is no longer an issue when compared to Chernikova-type error localization algorithms. This is very important since reducing computations is a critical aspect of developing a Fellegi-Holt system.

While doing this research we found that the balance equations may affect the ratio edits bounds and bounds in the complete set of ratio edits are not necessarily optimal. The following lemma tells us that if two details are required to balance to a reported total and two terms of this balance equation are in a ratio edit then we need to verify whether the lower or upper bounds for the ratio needs to be adjusted.

Lemma 1: If fields v_i and v_j balance to total v_t, $v_i + v_j = v_t$, then the bounds of the ratio edits connecting fields v_i, v_j, and v_t are not necessarily optimal and may need to be adjusted using the interaction with the balance equation.

Proof: For simplicity we consider only one case, all others follow similarly.
Let $v_i + v_j - v_t \leq 0$ and $v_i - u_{ij} v_j \leq 0$ be linear inequality edits corresponding to the balance equation and ratio edit respectively. Since the coefficients of v_j have opposite signs we can use Fourier-Motzkin elimination to generate a new induced edit. This new edit is a ratio edit, $\frac{v_i}{v_t} \leq \frac{u_{ij}}{1 + u_{ij}}$, since there are two fields in common in the generating balance equation and ratio edit.

In this case if $\frac{u_{ij}}{1 + u_{ij}} \leq u_{it}$, then we have found a more restrictive upper bound u_{it} for the ratio connecting fields v_i and v_t, therefore the upper bound is not optimal and needs to be adjusted.

Corollary: All ratio edit bounds are not necessarily optimal and may need to be adjusted due to the interaction with the balance equations.

The previous result follows from the fact that any pair of ratio edits with a common data field implies another ratio edit. Therefore, updating at least one bound in the complete set of edits implies that all lower and upper ratio edit bounds must be revised and updated. In the next section we will see that in our test data,136 ratio edits in 17 fields for each NAICS code, 15% of the lower and upper bounds were adjusted after two passes through the new implicit edit generation program. The possibility that the ratio edits bounds should be modified using the edit restrictions imposed on data items by the balance equations has not been considered in the earlier version of the SPEER edit system. It implies that the algorithms in the previous version of SPEER did not have available the edits that impose the most restrictions on the data fields, and therefore could change the error localization solutions and the imputation intervals used to "fill-in" data in the imputation algorithms.

The implicit edits generated by ratio edits and balance equations are computed using the methodology described above. The code is written in SAS and SAS/IML. The input of the new implicit edit generation module is the complete set of ratio edits and the balance equations. We first generate all implicit edits obtained by eliminating a common variable from a ratio edit and balance equation. The edit generation program then successively generates implicit edits by combining the newly generated implicit edits with the ratio edits. In their research, Draper and Winkler (1997) showed that this type of edits, obtained by replacing terms in a balance equation with the appropriate terms from the ratio edits, allows the SPEER system to error localize most edit failing records. This result is very important: it allows us to consider the smaller subset of the implied edits obtained by combining the newly generated edits with the ratio edits only which greatly simplifies the implicit edit generation methodology.

The algorithm used in the implicit edit generation is as follows:

> **Step 1.** Represent the ratio edits and balance equations as homogeneous linear inequality edits, $Av \leq 0$, $A = \left(\frac{R}{B}\right)$, where R and B are the matrices of coefficients corresponding to ratio and balance edits respectively, and v is the vector of data fields.
> **Step 2:** Choose two linear inequality edits with a common field v_k in which the coefficients of v_k have opposite signs in the ratio and balance edits. Use Fourier-Motzkin elimination to generate a new implied edit.
> **Step 3:** Verify that the new implied edit is a new derived edit. If the new implied edit has only two entering fields then check whether the corresponding ratio edit bound needs to be updated. If any ratio edit bound is updated then revise and update the complete set of ratio edits.
> **Step 4:** Adjoin the coefficients from the new implied edits to the matrix of coefficients A, and go to Step 2.

2.3 Editing in the new SPEER

The current version of SPEER (Draper and Winkler, 1997) for editing numeric data under ratio edits and single level balancing generates failing implicit edits during error localization for every edit failing record. In the previous section we described how the Fourier-Motzkin elimination method can be used to generate linear inequality edits implied by ratio and single-level balancing edits. In the new version of SPEER we use this methodology to generate a large subset of the implicit edits prior to automatic editing which considerably simplifies error localization in the SPEER edit system. This is important because the implicit edits are then available to be used many times in the error localization routine for every edit failing record. The need to repeatedly generate the implicit edits for every edit failing record is eliminated and the computational effort during error localization is reduced.

In the new version of SPEER, the edit checking, the error localization, and the imputation modules have all been rewritten to use the implicit edits generated prior to automatic editing. The edit checking routine identifies the records failing any ratio edit, balance equation, or implicit edit. Changes to the edit checking routine are straightforward, we simply added code to determine if any of the implicit edits generated using the new algorithm failed. The code in the previous version of the error localization module needed to generate and error localize failing implied edits was not particularly easy, and it is no longer needed. Error localization has been greatly simplified. For every data record marked as failing at least one edit (ratio or balance) in the edit checking routine, the error localization module uses a greedy algorithm (Nemhauser and Wolsey 1987) to determine the minimum number of fields to impute so that the record no longer fails.

The code in the imputation algorithm also uses the information from the implicit edits generated prior to automatic editing. We recall that one of the main results of the Fellegi-Holt (Fellegi and Holt, 1976) theory is that if we know the values of a subset of fields that satisfy all edits that place restrictions on those fields only, then we can impute for the remaining fields so that the record satisfies all edits. The imputation routine will successively check each field identified to be changed and impute for that item. If there is only one term in a balance equation marked for imputation, then the balance equation is used to impute the value of the item. Otherwise, we impute a field value using the information from the other known fields' values, the ratio edits restrictions, balance edits and implied edits to determine the interval into which to impute. Draper and Winkler (1997) showed that the implied edits generated by a failing ratio edit and a balance equation are sufficient for determining the imputation intervals. We used this result in the new imputation routine by using only the implied edits generated the first time through Step 2 in the edit generation algorithm described in Section 2.2.

The algorithm for SPEER editing is as follows:

> **Step 1:** For each record, use ratio edits and balance equations to identify edit failures. If record fails at least one edit, use induced edits generated using the methodology described in Section 2.2 to identify failing induced edits. Otherwise, go to the next record.
> **Step 2:** Use the failing ratio edits, failing balance equations, and failing implicit edits identified in Step 1 in a greedy algorithm to determine the number of fields to be changed so that the record satisfies the edits.
> **Step 3:** For each field marked to be imputed in Step 2, determine if the item value can be imputed using a balance equation. Otherwise, use the other known fields (reported and imputed), the ratio and balance edits, and the first order induced edits to determine an interval into which field values can be imputed.

Section 3: Results

To test the new SPEER'02 algorithms we used keyed data from the 1997 Annual Survey of Manufactures (ASM). The ASM collects data from manufacturing establishments on a four page paper instrument. The ASM measures manufacturing activity that includes employment, payroll, fringe benefits, cost of materials, product shipments, capital expenditures, and total inventories. The ASM also provides measures of industrial production and productivity. This survey is the only source of comprehensive data on the manufacturing level of the USA economy.

Our test data consists of 6,533 records on 310 industry classification codes (NAICS). Each record contains an identification number, a NAICS code, and data for 17 numerical fields. The ASM fields edited using the SPEER editing system are listed in Table 1. ASM fields measuring production worker wages (WW) and other employee wages (OW) are required to balance to the reporting unit's total salary and wages (SW). Similarly the number of production workers (PW) and other employees (OW) must be equal to the reported total employment (TE). The last four fields (PTIE, PTIB, PVS, PCM) contain the calculated sum of detail items corresponding to their respective totals.

Table 1: ASM Data Fields to be Edited in SPEER

ASM Fields	Description
SW = WW + OW	Salary and Wages
VS	Value of Shipments
TE = PW + OE	Total Employment
WW	Production Worker Wages
OW	Other Employee Wages
TIB	Total Inventory at Beginning of Year
CM	Cost of Materials
TIE	Total Inventory at End of Year
PW	Number of Production Workers
OE	Number of Other Employees
PH	Number of Plant Hours Worked
LE	Legally Required Benefits
VP	Voluntarily Paid Fringe Benefits
PTIE	Calculated Sum of Details of TIE
PTIB	Calculated Sum of Details of TIB
PVS	Calculated Sum of Details of VS
PCM	Calculated Sum of Details of CM

The explicit ratio edits are defined by the subject matter experts. The auxiliary program for implicit ratio edit generation is used to generate ratio edit bounds for every pair of fields. In our test data there are 310 industry classes, 136 ratios for each class, for a total of 84,320 linear inequality edits corresponding to the complete set of ratio edits.

The complete set of edits and the balance equations are then used as input to the implicit edit generation program. In the previous section we mentioned that it was possible that the ratio edit bounds needed to be adjusted during implicit edit generation –this is important since the ratio edits bounds are used for computing imputation intervals so that record no longer fails. Table 2 displays the total number of ratio edit bounds adjusted after two passes through the implicit edit generation program. For the ASM edits, 15% of the ratio edit bounds were adjusted after two passes through the implicit edit generation program.

Table 2: Number of Adjusted Ratio Edit Bounds in the ASM Ratio Edits

Number of Items in ASM Data	Number of NAICS in Test Data	Number of Ratio Edits for each NAICS code	Number of Bounds Adjusted After Two Passes
17	310	136 (272 bounds)	12,346 (15%) (out of 84,320)

The set of linear inequality edits generated using the implicit edit generation program is used, along with the adjusted complete set of ratio edits, as input to the new SPEER system. We used the test data of 6,533 1997 ASM records described above for comparing the results when running the 1997 version of SPEER (SPEER' 97) and the new version of SPEER (SPEER' 02). We examined how many records can be automatically corrected by either system so that an edit failing record no longer fails after doing multiple passes through the data. After the first pass through the editing system, imputation will not be successful for a small proportion of records. These records will be partially corrected by imputing only those fields for which the imputation routine successfully computed imputation intervals. These partially corrected records are then passed again through the editing system.

In our test data, both programs identified all 6,533 records as edit failing records, however the number of records corrected by either program after different passes through the data is different. Table 3 displays the number of records still failing edits after different passes through the editing system. Clearly both edit systems are performing quite well in terms of correctly identifying items to impute so records no longer fail. There are 297 records still failing edits after one pass through SPEER’97, while the number of records still failing edits in SPEER’02 is 104. The number of records still failing edits after two passes is 57 (0.9%) in SPEER’02 and 81 (1.2%) in SPEER’97. SPEER’02 consistently corrects more records in the first and second passes than SPEER’97 –more than 99% of the records are corrected in these two passes and there is no significant gain in records corrected after running the data through the system a third time.

Table 3: Number of Records Still Failing After Different Passes Through SPEER’02 and SPEER’97

Pass	SPEER’97	SPEER’02
First Pass	297 (4.5%)	104 (1.6%)
Second Pass	81 (1.2%)	57 (0.9%)
Third Pass	42 (0.6%)	54 (0.8%)

Our next comparison examines the effect of using the large subset of the implied edits generated a priori on the number of times a field was marked for deletion during error localization. We recall that SPEER’97 calculates failing implicit edits generated by failing ratios and balance equations only if needed, while SPEER’02 uses the large subset of the implicit edits generated prior to error localization. Table 4 displays the ASM fields identified to be imputed and the number of times each field was marked to be changed for a subset of the data including only records for which all fields marked for deletion were successfully imputed after two passes through the data (6,408 records). The number of times reported details WW, OW, and reported total SW are marked to be changed is larger in SPEER'02 than in SPEER'97. This result is expected. In SPEER'02, fields restricted by a balance edit enter all the implicit edits generated prior to error localization: the error localization module uses a greedy algorithm, therefore the number of times a field is marked for deletion is associated to the number of times the field enters the failing edits. For all other fields (with the one exception of item PH), the number of times a field was marked for deletion during error localization is consistently higher in SPEER'97 when compare with SPEER'02.

Table 4: Number of Times Field was Marked for Deletion After Two Passes in SPEER'97 and SPEER'02

ASM fields	Number of times field was changed in SPEER' 97	Number of times field was changed in SPEER' 02
SW = WW + OW	453	2515
VS	447	436
TE = PW + OE	418	357
WW	181	2309
OW	428	483
TIB	31	29
CM	105	93
TIE	27	26
PW	556	372
OE	747	443
PH	608	624
LE	554	508
VP	245	218
PTIE	280	279
PTIB	260	258
PVS	3448	3447
PCM	368	356

4. Discussion

In this paper we described a new implicit edit generation algorithm for the SPEER edit system based on the Fourier-Motzkin methodology for finding solutions to a system of linear inequality edits. The system takes as input the complete set of ratio edits and the balance equations. The set of ratio edits and balance equations are then represented as linear inequality edits. These linear inequality edits are then used to generate a subset of the implicit edits. The implicit edits that are generated are checked and any redundant edits are discarded. The software has an option for choosing the maximum number of passes through the system.

This paper presented theory, algorithms and results from testing the new version of SPEER algorithms on a subset of edit failing records from the Annual Survey of Manufactures production data. The new version of SPEER is exceptionally fast –the system error localized and successfully imputed 99% of the records (6,533 records, all edit failing records) in two passes through the data in 90 seconds (clock time, about 66 records per second.) Using this methodology has several potential advantages for Census Bureau's SPEER editing system. First, the logic needed to implement the algorithm for the edit generation system and SPEER editing are simple, easy to understand and can be used with any survey under ratio edits and single level balancing. Using this new algorithm has the added advantage that the implicit edits are generated once, prior to SPEER editing, and are available to be used repeatedly during error localization for every edit failing record. This greatly simplifies the code in the error localization module since there is no need to generate failing implicit edits for every edit failing record. This approach is not however without its disadvantage: generating a large subset of implicit edits for some surveys could possibly take considerable time and the set of implicit edits can grow very large.

4. References

DeWaal, T., (1996), "CherryPi: A computer program for automatic edit and imputation," *UN Work Session on Statistical Data Editing*, November 1996, Voorburg.

Draper, L., and Winkler, W., (1997), "Balancing and Ratio Editing with the New SPEER System," *American Statistical Association, Proceedings of the 1997 Section on Survey Research Methods*, 582-587.

Duffin, R. J., (1974), "On Fourier's Analysis of Linear Inequality Systems," *Mathematical Programming Study*, North-Holland Publishing Company, 71-93.

Fellegi, I. P. and D. Holt, (1976), "A Systematic Approach to Automatic Edit and Imputation," *Journal of the American Statistical association*, 71, 17-35.

Greenberg, B. and Surdi, R., (1984), "A Flexible and Interactive Edit and Imputation System for Ratio Edits," SRD report RR-84/18, US Bureau of the Census, Washington, D.C.

Greenberg, B. and Petkunas, T., (1990), "SPEER (Structured Programs for Economic Editing and Referrals), *American Statistical Association, Proceedings of the 1990 Section on Survey Research Methods.*

Nemhauser, G. L. and Wolsey, L. A., (1988), "*Integer and Combinatorial Optimization*,"John Wiley: NY.

Quere, R., (2000), "Automatic Editing of Numerical Data," *Report, Statistics Netherlands*, Voorburg.

Rubin, D. S., (1975), "Vertex Generation in cardinality Constrained Linear Programs," *Operations Research*, No. 23, 555-565.

Sande, G., (1978), "An Algorithm for the fields to Impute Problems of Numerical and Coded Data," Technical Report, Statistics Canada.

Schiopu-Kratina, I. and Kovar, J. (1989), "Use of Chernikova's Algorithm in the Generalized Edit and Imputation System," Statistics Canada, Methodology Branch Working Paper BSMD 89-001E.

Thompson, K. J. and Sigman, R. (1996), "Statistical Methods for Developing Ratio Edits Tolerances for Economic Censuses," *American Statistical Association*, *Proceedings of the Section on Survey Research Methods*.

Todaro, T. (1999). "Evaluation of the AGGIES Automated Edit and Imputation System," National Agricultural Statistics Service, USDA, Washington D. C., RD Research Report No. RD-99-01.

Winkler, W., (1997) "Set Covering and Editing Discrete Data," *American Statistical Association*, *Proceedings of the Section on Survey Research Methods.*

Winkler, W. and Chen, B. C. (2002), "Extending the Fellegi-Holt Model of Statistical Data Editing," *Research Report Series Statistics* RRS2002/02, Statistical Research Division, US Census Bureau, Washington D.C.