

# USING AN EDIT SYSTEM TO DEVELOP EDITING SPECIFICATIONS

Brian Greenberg, U.S. Bureau of the Census

## Abstract

All survey and census data, economic data in particular, must be checked for consistency and reasonableness. For each respondent in an economic survey or census, data items such as value of shipments, salary and wages, and total employment are compared against each other by means of a variety of arithmetic and logical tests to determine if one or more has aberrant values. A widely used edit criterion is the ratio edit which requires that the ratio of two selected response items lie between specified bounds. The upper and lower bounds are usually established by analyzing historical evidence, by drawing upon subject-matter expertise, and, when feasible, by examining a sample of responses. An edit system having proper diagnostics, facilities to monitor performance, and the ability to expeditiously alter specifications, generate logical consequences of edits, and detect inconsistencies can be a valuable tool in determining requisite bounds. Such an edit system has been developed at the Bureau of the Census for the editing of economic data.

## I. INTRODUCTION

All survey and census data, economic data in particular, must be checked for consistency and reasonableness. For each respondent, the various fields, such as value of shipments, salary and wages, and total employment, are compared against each other in a variety of ratio edits to determine if one or more of them have aberrant values. For example, evidence may indicate that the ratio of salary and wages to total number of employees in a particular industry (i.e., the average income per employee for that industry) usually lies between two prescribed bounds. In addition, it is expected that the ratio of salary and wages to value of shipments for a given respondent will not fluctuate drastically from one year to the next. In fact, evidence indicates that the value of that particular ratio for the current year response divided by the value of that ratio for the prior year recorded response will also lie between two prescribed limits. The first requirement above is referred to as a **field-to-field ratio edit** and the second as a **year-to-year ratio edit**.

In general, the objective of most automated edit and imputation systems is: given the response to a survey or census (and the respondent's prior year recorded response or other ancillary information) either to verify that the current year record passes all edits or to locate a set of fields to be adjusted in the current year response so that all edits can be passed. The edit and imputation system then allocates values to those fields not reported and alters the responses in the fields targeted for change due to some pattern of edit failure. The system then creates a revised record that is consistent with respect to the full edit set (except for deliberate exceptions) and, when appropriate, refers records for analyst review.

The process of locating a set of fields in which values should be altered in order that the adjusted record will pass all appropriate edits is referred to as **error localization**. A **revised record** is then created by altering the values in the fields targeted for change and by allocating values in the fields that were not

reported. **Imputation** is the assignment of a value to a field either for non-response or to replace a recorded value determined to be inconsistent with a set of edits.

The questions we will address in this paper are: (1) what is the acceptable range for imputed values so that all ratio tests are satisfied? (2) how does one determine the bounds for ratio edits? A well structured edit and imputation system that is easy to work with and provides ample diagnostic and summary statistics can be a valuable tool for subject-matter specialists in dealing with these questions. We discuss an edit and imputation system developed in the Statistical Research Division of the Bureau of the Census that addresses the questions above and relate experience in using it. A prototype of this edit system using data and edits taken from the Annual Survey of Manufactures is now being tested. In addition, this system has been employed to test the consistency of the user-provided ratio edits, to derive the logical implications of these edits, and to aid in determining suitable upper and lower ratio limits.

Gordon Sande has addressed the question of using an edit and imputation system to analyze editing criteria when the edits are linear inequalities. He combines the use of summary statistics and analyses based on linear programming. A discussion of his methods is contained in [3].

In this paper we only consider the case of field-to-field edits. Similar considerations can be made to apply to year-to-year edits as well.

## II. THE CORE EDIT—OVERVIEW

### Section 1. Basic Goals

In this section we will discuss a mathematically-based, structured editing system, referred to as the **core edit**. The **core edit** will be incorporated into a larger system for the editing and correction of economic data under ratio tests. Provided with the recorded response to a survey or census, the **core edit** either verifies that the current year record passes all edits or it locates a (weighted) minimal set of fields to be adjusted in the current year response. The **core edit** then allocates values to those fields not reported and alters the responses in the fields targeted for change, creating a revised record that is consistent with respect to the full edit set or, in some circumstances, a subset of the full edit set.

### Section 2. Structure

The **core edit** is organized into three major segments: the EDIT CHECKING subroutine, the ERROR LOCALIZATION subroutine, and the IMPUTATION subroutine. These subroutines are embodied in a main driving program. A record and, when available, its associated prior year record and any other ancillary information are read by the main program and passed to the EDIT CHECKING subroutine. A list of edit failures is compiled and passed to the ERROR LOCALIZATION subroutine which determines fields to alter (if any) based on the pattern of edit failures and field weighting criteria. The list of fields to be changed, the reported values, the prior year record, and other ancillary information are passed to the IMPUTATION subroutine. The fields

targeted for change and fields not reported are sequentially imputed. Values reported in unchanged fields or imputed in an earlier stage are used to determine an acceptable range (the **acceptance region**) for values in fields yet to be imputed. A revised record is created and returned to the main program. For each field to be imputed a separate **imputation module** is called which incorporates imputation strategy for that particular field. These imputation modules are determined with the aid of subject-matter personnel and reflect their understanding of the field being imputed and the underlying survey environment.

### III. EDIT GENERATION AND EDIT ANALYSIS

#### Section 1. User-Supplied Edits

Let there be  $N$  fields,  $F_1, \dots, F_N$ , and let  $A(i)$  the current year values in field  $F_i$  for a given respondent.

A field-to-field ratio test involving fields  $F_k$  and  $F_h$  is the requirement that

$$L_{kh} \leq A(k)/A(h) \leq U_{kh}.$$

Each ratio gives rise to two edits, namely:

$$L_{kh}A(h) - A(k) \leq 0$$

$$A(k) - U_{kh}A(h) \leq 0.$$

One can view the original edit set as a graph in which the fields correspond to nodes, and an edit between two fields to an arc between the corresponding nodes. If this graph is connected, when we generate the implied edits we, in essence, complete the graph. If the original graph is not connected, we work with the connected components and, in effect, partition the set of fields into subsets each of which is edited independently of the other. We will assume, therefore, that any two fields can be linked by a path of ratio edits.

#### Section 2. Implied Edits

In their paper [1], Fellegi and Holt describe a procedure wherein, given a set of edits, one generates a family of **implied edits** (i.e., logical consequences of the given edits), which, when used with a set covering procedure on the subsets of fields that mutually fail an edit, enables one to locate the minimal number of fields to impute in a given record. This procedure, as a rule, is not practical because the number of implied edits grows quite large. When working with ratio edits, however, it is not difficult to get a complete set of implied edits. Given the ratio tests

$$L_{kh} \leq A(k)/A(h) \leq U_{kh}$$

$$L_{hj} \leq A(h)/A(j) \leq U_{hj}$$

the implied ratio inequality is

$$L_{kh}L_{hj} \leq A(k)/A(j) \leq U_{kh}U_{hj}.$$

This implied ratio test gives rise to two implied edits, and through an iterative process a complete set of implied edits arises in this fashion. One can see that there are at most  $2\binom{N}{2} = N(N-1)$  edits in the complete edit set. It is at this stage inconsistencies in the edit set will surface (i.e., a lower bound for some ratio will exceed its upper bound), and accordingly, any inconsistencies unearthed by the implied edits require

review by subject-matter specialists.

In addition to deriving the implied edits relating two fields that were not involved in a single user-supplied edit, the edit analysis also can narrow the limits on user-supplied ratios. For example, a user-supplied edit may be:

$$L'_{kh} \leq A(k)/A(h) \leq U'_{kh}.$$

By employing the process discussed above, an implied edit might be:

$$L''_{kh} \leq A(k)/A(h) \leq U''_{kh}.$$

The edit to be used relating fields  $F_k$  and  $F_h$  in the EDIT CHECKING subroutine will be

$$L_{kh} \leq A(k)/A(h) \leq U_{kh}$$

where:

$$L_{kh} = \max \{ L'_{kh}, L''_{kh} \} \quad \text{and}$$

$$U_{kh} = \min \{ U'_{kh}, U''_{kh} \}.$$

In the EDIT CHECKING routine all fields are compared against one another by using the implied edits. That is, for every pair of fields,  $F_k$  and  $F_h$ , we say the pair  $(F_k, F_h)$  is involved in a failed field-to-field edit if either

$$A(k)/A(h) < L_{kh} \quad \text{or}$$

$$A(k)/A(h) > U_{kh}.$$

We compile a list of pairs of fields involved in a failed field-to-field edit or a failed year-to-year edit and pass them to the ERROR LOCALIZATION subroutine.

#### Section 3. The EDIT ANALYZER

The purpose of the EDIT ANALYZER is to derive and exhibit the logical implications of the user-provided edits and to detect any inconsistencies among them. After all ratio edits and associated bounds have been read in, the system generates the implied edits. The algorithm incorporated in the system is basically the one indicated by Fellegi and Holt in [1]. For every pair of fields, say  $F_k$  and  $F_h$ , we get an implied ratio edit

$$L_{kh} \leq A(k)/A(h) \leq U_{kh}.$$

These ratios are all listed. If there is an inconsistency in the edit set, the message that an inconsistency has been detected is printed along with some indication of the edits responsible. In some cases the lower limit will be zero and/or the upper limit will be infinite.

#### Section 4. Two Examples

Example 1:

As an example of a set of edits and the attendant implied edits, suppose we have a set of records each having only three fields:  $F_1$ ,  $F_2$ , and  $F_3$ . Suppose also that the user-supplied edits are

$$(H1) \quad 2 \leq A(1)/A(2) \leq 4$$

$$(H2) \quad 1/2 \leq A(2)/A(3) \leq 1.$$

An implied edit is:

$$(H3) \quad 1 \leq A(1)/A(3) \leq 4.$$

That is, we make explicit the implied relation between

fields  $F_1$  and  $F_3$ .

Suppose further we have the record  $A = (4,3,9)$ , (that is,  $A(1) = 4$ ,  $A(2) = 3$ , and  $A(3) = 9$ ). Note that (H1) and (H2) are both failed by that record, and it might seem at first glance that field  $F_2$ , the only field involved in both (H1) and (H2), is the culprit. But one also observes that (H3) is failed by the record so that we will have to change at least two fields in the record  $A$  for it to pass all edits. Observe, incidentally, that if one attempted to alter only  $F_2$ , no value for  $A(2)$  would pass (H1) and (H2) given the values for  $A(1)$  and  $A(3)$ , so a consistent revised record could not be created. When there are many more fields than just three, the importance of the implied edits becomes even more apparent.

#### Example 2:

As discussed, the logical implications of the user supplied edits not only make explicit certain relations that were only implicit before, but also affect other edits furnished by subject-matter personnel. For example, suppose three edits are supplied by the user

- (E1)  $3 \leq A(1)/A(2) \leq 6$
- (E2)  $2 \leq A(2)/A(3) \leq 4$
- (E3)  $4 \leq A(1)/A(3) \leq 12$ .

By combining (E1), (E2), and E(3) as indicated in equations (3) and (4), we obtain the implied edit

$$(E4) \quad 6 \leq A(1)/A(3) \leq 12.$$

Note that the lower limit was obtained jointly from (E1) and (E2), and the upper limit was taken directly from (E3).

It is then up to the subject-matter specialist to determine the preferred editing criterion involving fields  $F_1$  and  $F_3$ . If, based on knowledge about the underlying environment from which the data is drawn, a ratio of  $F_1$  to  $F_3$  may quite reasonably be, say  $\frac{5}{2}$ , then it is incumbent that the lower bound of either (E1) or (E2) be altered to reflect that fact. If, on the other hand, subject-matter specialists feel strongly that (E1) and (E2) should remain as they are, then the lower bound of (E3) should be adjusted or replaced by that of (E4) to reflect the implication of (E1) and (E2).

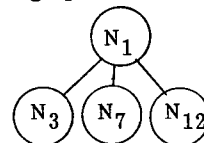
In the EDIT CHECKING routine of the **core edit**, edit (E4) will be used to check for edit failures, and a message that an edit is failed between fields  $F_1$  and  $F_3$  will be based on edit (E4) rather than edit (E3).

## IV. ERROR LOCALIZATION AND IMPUTATION STRATEGY

### Section 1. Error Localization

After the set of edit failures is determined, we select the fields to target for adjustment in the ERROR LOCALIZATION routine. A graph is created that captures the pattern of edit failures where the nodes correspond to fields. Node  $N_i$  corresponds to field  $F_i$  where  $i = 1, \dots, N$ , and the weight of node  $N_i$  is defined to be the weight of the corresponding field. The set of edit failures passed from the EDIT CHECKING routine are used to create the **failed edit graph**. For each failed record we form the **failed edit graph** in which the nodes correspond to fields failing at least one edit, and an arc between two nodes indicates that the corresponding fields fail an edit between them. For example, if the only edit failures in a record involve fields  $F_1$  and  $F_3$ ,  $F_1$  and  $F_7$ , and  $F_1$  and

$F_{12}$  the **failed edit graph** would be:



The **failed edit graph** is disconnected by removing one node at a time and by removing all arcs incident with a deleted node. When no arcs are left, the graph is said to be **totally disconnected**, and we stop. The union of the set of fields corresponding to nodes remaining and those nodes not in the graph in the first place are mutually consistent with respect to our edit set since no two of them are involved in a failed edit (since there are no arcs remaining). If we assign values to all deleted fields in such a way that no edit failures are engendered, we will have created a consistent revised record. Because the non-deleted fields are mutually consistent, whenever we assume the edits are consistent we can make imputations in fields deleted from the **failed edit graph** in such a manner that the adjusted record passes all edits, see [1,2] for more details.

Accordingly, by selecting a set of nodes to delete in disconnecting the failed edit graph, we have chosen a set of fields to impute. We associate a low weight to a field we are reluctant to change, and target for alteration fields with a high weight and/or a large number of edit failures. For each node in the graph we construct the **weighted degree** (that is, the number of arcs into a node multiplied by the weight of that node). At present, we employ a procedure which removes, one at a time, nodes of high weighted degree or nodes adjacent to nodes of low weighted degree. In so doing, we disconnect the failed edit graph (and, thus, provide a set of fields to delete and impute for).

### Section 2. The Acceptance Region

Suppose a set of fields has been selected for imputation either due to edit failures or because of non-response. By a suitable renumbering, we can assume that fields  $F_1, F_2, \dots, F_{k-1}$ ,  $k \leq N$  have valid values and that we want to impute field  $F_k$ . For each field  $j < k$ , we have an edit between field  $F_j$  and field  $F_k$  due to the generation of implied edits. That is, for each  $j < k$  we have an inequality:

$$L_{kj} \leq A(k)/A(j) \leq U_{kj}$$

By multiplying by  $A(j)$ , we obtain, for each  $j < k$ , a family of linear inequalities:

$$(M_j): \quad L_{kj}A(j) \leq A(k) \leq U_{kj}A(j).$$

Noting that  $L_{kj}$ ,  $U_{kj}$ , and  $A(j)$  are known real numbers, each  $M_j$  is an interval on the real line. The intersection of the  $M_j$  for all  $j < k$ , call it  $I_k$ , is the set of points

$$\{x \mid L_k \leq x \leq U_k\} \quad \text{where}$$

$$L_k = \max \{A(j)L_{kj} \mid j < k\}$$

$$U_k = \min \{A(j)U_{kj} \mid j < k\}.$$

We call  $I_k$  the **acceptance region** for field  $F_k$  and note that  $I_k$  is non-empty because the set of field-to-field edits is assumed consistent. Any number in  $I_k$  if imputed for  $F_k$ , will be consistent with all values  $A(j)$ ,

$j < k$ . That is, if

$$L_k \leq A(k) \leq U_k$$

the set of fields  $F_1, \dots, F_k$  will be mutually consistent, and conversely.

If  $k+1 \leq N$ , repeat the procedure above for field  $F_{k+1}$ , and so on, until all fields are imputed. At each juncture, we employ all valid reported responses and all values imputed at an earlier stage to obtain the **acceptance region** for the field to be imputed. After as many iterations as there are fields to be imputed, we have created a complete revised record, all of whose values are mutually consistent.

### Section 3. An Example

Example 3:

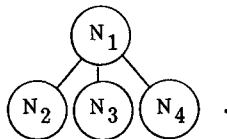
The following is an example of the entire process as described so far. Let  $A = (4, 4, 9, 17)$  and let the user-provided edits be:

$$\begin{aligned} (G1) \quad & 2 \leq A(1)/A(2) \leq 4 \\ (G2) \quad & 1/4 \leq A(2)/A(3) \leq 1 \\ (G3) \quad & 1/2 \leq A(3)/A(4) \leq 1. \end{aligned}$$

The implied edits are:

$$\begin{aligned} (G4) \quad & 1/2 \leq A(1)/A(3) \leq 4 \\ (G5) \quad & 1/8 \leq A(2)/A(4) \leq 1 \\ (G6) \quad & 1/4 \leq A(1)/A(4) \leq 4. \end{aligned}$$

The edits that fail for  $A$  are (G1), (G4), and (G6), and the failed edit graph is:



Clearly field  $F_1$  is the field to be adjusted. Accordingly we obtain:

$$\begin{aligned} 2 A(2) &\leq A(1) \leq 4 A(2) && \text{from (G1)} \\ 1/2 A(3) &\leq A(1) \leq 4 A(3) && \text{from (G4)} \\ 1/4 A(4) &\leq A(1) \leq 4 A(4) && \text{from (G6)}. \end{aligned}$$

Since  $A = (4, 4, 9, 17)$ , we have:

$$\begin{aligned} 8 &\leq A(1) \leq 16, \\ 4.5 &\leq A(1) \leq 36, && \text{and} \\ 4.25 &\leq A(1) \leq 68. \end{aligned}$$

Thus,  $8 \leq A(1) \leq 16$ ,

and any valid imputation for  $F_1$  must lie within this range (the **acceptance region**).

### V. DIAGNOSTIC INFORMATION FROM THE CORE EDIT

In the running of a deck of records, information concerning the performance of the system is made available in two ways: as part of the output for each record which fails one or more edits and as a set of summary statistics after the entire deck is run.

As each failed record is processed through the **core edit**, the following information is printed out. First, the record itself, followed by the adjustments due to simple balancing. Following that is a list of the

failed field-to-field edits, then a list of the failed year-to-year edits. (The list includes any failed implied edits.) Next comes a list of fields not reported and a list of fields targeted for change due to edit failures. The **revised record** is printed next. For each field imputed, along with the adjusted field values we record the lower and the upper limits of the range in which the adjusted field value must lie. An analyst reviewing the record can choose any value in the acceptable range for imputation and still create an **acceptable revision**.

Selected messages are printed at the end of each record run: whether or not the revised record fails any edits, if either the field-to-field or the year-to-year edits are bypassed, if the revised record is an **acceptable revision**, if it was referred for analyst review, if a successful imputation was impossible due to some inconsistency in the edits, etc. By a careful reading of the system messages, a person reviewing the processing of any record can garner a considerable amount of information, understand the rationale for each system decision, and make an informed determination for alternative action.

After an entire deck of records is run the following information is printed as summary information. For **each edit** (implied as well as user-supplied) we list the number of times it is failed throughout the deck, and for each user-supplied edit we list the identifying sequence number of the records which failed it. For **each field**, we provide the number of times it is not reported and the number of times it is adjusted due to edit failures. The objective of these diagnostics is to enable a user to evaluate the impact of the editing criteria on the data with the aim of making informed adjustments to the edits if needed.

### VI. DEVELOPING A PROTOTYPE OF THE SYSTEM

As stated earlier, a prototype version of the **core edit** has been developed based on the Annual Survey of Manufactures (ASM). We integrated the various segments of the system that were described above and incorporated the expertise of Industry Division subject-matter personnel in creating the **imputation modules** and special purpose functions. In addition to creating a workable prototype, we wanted the survey specific routines and functions to reside in modules that could be altered with facility in order to enable one to update, modify, and revise the system to suit changing needs. Furthermore, we want to employ the ASM prototype in editing other surveys by changing only survey-specific segments.

Subject-matter personnel selected twenty fields to use in the **core edit** and provided twenty-eight field-to-field and eighteen year-to-year ratio edits involving these fields. The edits were run through the EDIT ANALYZER and no inconsistencies were detected. Although a number of the implied edits proved to be quite useful and revealing for purposes of analysis, many were edits between fields that were only weakly related. In the later case, these edits provided little useful information.

We obtained a sample of data from three industry groups which included both the current year record and prior year record, and ran the data through the EDIT CHECKING and ERROR LOCALIZATION routines. We observed which fields had excessively high failure rates and which were frequently targeted for change. We also observed which edits were being failed excessively often. For the edits that appeared to be

failing too frequently, we went back and examined the records which failed these edits. Based on an examination of the records themselves, we made several changes in the original edits, for the most part relaxing the tolerance range. The entire edit set was once again run through the EDIT ANALYZER and no inconsistencies were present.

We next created **imputation modules** for each of the twenty fields. The options and strategy for each **imputation module** were created with the aid of subject-matter personnel working on the project. They tried, in part, to simulate the decision process that would be followed by an analyst examining the record. Typically, for each field, the associated **imputation module** had about five or six options and consisted of about fifty lines of fairly simple Fortran code.

## VII. DEVELOPING EDITING CRITERIA

A second study was undertaken with the cooperation of the Enterprise Statistics Branch, Economic Surveys Division. The first objective of the study was to test the practicality and effectiveness of the edit system as an aid in the development of a set of edit specifications. The second was to determine the facility with which a prototype of the edit system, created for the Annual Survey of Manufactures (ASM), could be altered to accommodate a new editing environment, in particular, one having no prior year data.

Staff from the Enterprise Statistics Branch selected twelve fields from the 1977 Company Summary Report Survey (CSRC) and twenty-eight field-to-field ratio edits relating these fields. As a starting point for the study, lower and upper ratio edit bounds were based on a sample of records in their file and assumptions about the respondent model. This starting set of ratio edit limits will be referred to as the "user-supplied" limits.

These edits were run through the EDIT ANALYZER and no inconsistencies were detected. Although many of the implied edits did not yield much information, some were quite useful. For some ratio edits a derived implied lower limit was greater than the lower limit on the "user-supplied" edits — the situation encountered in **Example 2**. That would have caused some ratios of selected fields to fail even though area specialists considered such a ratio of responses to be acceptable. Accordingly, a number of changes were made to some user-supplied limits so that a derived lower limit would not exceed an intended value. The need for these adjustments would not be as apparent with a less systematic, albeit careful, examination of the original edit set. The revised set of edits was once again run through the EDIT ANALYZER, no inconsistencies were detected, and we proceeded to the next stage of the program.

A sample of response data was run through the EDIT CHECKING and ERROR LOCALIZATION subroutines of the main program. After the run was completed, summary information was displayed. The number of times each edit was failed was listed (this included the implied edits). For each user-supplied edit the records which failed it were also listed. Edits that were being failed excessively often were suspicious. By listing the records on which these failures occurred, the subject-matter personnel were able to evaluate whether or not the rejections were justified, that is, whether a pair of field values that

both looked reasonable were failing an edit. Based on this analysis, the user-supplied limits were modified.

In addition, the number of times each field failed an edit was displayed along with the number of times each field was targeted for change. If a given field was involved in an excessive number of edit failures it might indicate that (1) it is subject to poor reporting, (2) it is a good discriminator of invalid responses in other fields, or (3) the edits involving that field are too restrictive. Based on an examination of the system output and focusing on the concerns above, a few additional changes were made to the user-supplied limits.

We next ran the records through the EDIT CHECKING, ERROR LOCALIZATION, and IMPUTATION subroutines. For the purpose of this study the **acceptance region** was of considerable interest. This range is determined by both the lower and upper limits on the ratio edits and by the values in fields considered valid. By exhibiting the **acceptance region** one can observe the effects of the editing parameters on the range of acceptable field values. By an evaluation of this range informed choices can be made concerning how (and if) to revise editing parameters. In general, if the **acceptance region** is very small, the edits may be too restrictive (in fact, if edits are inconsistent, the **acceptance region** is empty). If the **acceptance region** is very large, the edits may be too lax (in fact, if the edits are totally non-discriminating, the **acceptance region** is unbounded).

In addition to examining the size of the **acceptance region**, it was of interest to observe how far outside the **acceptance region** a deleted reported value fell. If a reported value fell only marginally outside, that would suggest that perhaps it might be accurate and should be considered an outlier rather than an invalid response. If for a given field, reported values altered due to edit failures frequently fell only slightly outside the **acceptance region**, the **acceptance region** should be expanded by extending the editing parameters involving that field. Because the **acceptance region** is determined, in part, by the editing parameters, the relation between the reported data and the **acceptance region** can provide information concerning the editing parameters. If a determination is made to expand (reduce) the **acceptance region**, then the lower and/or upper ratio limits would have to be extended (contracted).

As it turns out, we found that a number of fields were being targeted for adjustment when their reported values were only slightly outside the derived **acceptance region**. Based on these observations and mindful of our overriding objective to keep from adjusting outliers (as opposed to erroneous responses), the lower and upper edit ratio limits employed as a starting point in our study (and already modified as noted above) were extended further. This newly revised edit set was run through the EDIT ANALYZER and found to be consistent. The data sample was then run against these revised ratio edits, and the edit performance was more satisfactory.

Thus, starting with candidates for edit criteria, the edit system was utilized to enhance and refine them and to make them reflect, to a greater degree, the performance desired by area specialists. At this stage, we brought the study to a close. We feel that the edit and imputation system discussed here (and certainly others as well) can have a role in developing and refining editing specifications.

The second objective in this study was to determine the portability of the ASM prototype. After altering input specifications and deleting references to prior year data and special purpose ASM routines, the system ran successfully on the CSRS. The modularized nature of the system along with the relatively uncomplicated programming involved, facilitated incorporating the changes required.

#### VIII. ACKNOWLEDGMENTS

The author would like to thank Paul Pietal formerly with Industry Division and Bob Rosati currently with Industry Division, Bureau of the Census for their assistance in the development of the ASM prototype of the **core edit**. We also thank John Monaco and Lorraine Tischler with Economic Surveys Division, Bureau of the Census for their work in utilizing the edit system to enhance and refine edit criteria.

#### REFERENCES

- [1] Fellegi, I.P. and Holt, D. (1976). A systematic approach to automatic edit and imputation. Journal of the American Statistical Association, 71, 17-35.
- [2] Greenberg, B. (1981). Developing an edit system for industry statistics. Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface, 11-16, Springer-Verlag, New York.
- [3] Sande, G. (1979). Numerical Edit and Imputation. International Association for Statistical Computing, 42nd Session of the International Statistical Institute.