# Measuring Edit Efficiency in the Economic Directorate of the U.S. Census Bureau [1]

Broderick E. Oliver and Katherine J. Thompson

United States Bureau of the Census
4600 Silver Hill Road, Washington, DC 20233
Broderick.e.oliver@census.gov
Katherine.j.Thompson@census.gov

## Introduction

In 2006-2007, the Economic Directorate of the United States Census Bureau conducted a series of studies to assess processing procedures of several of its Economics surveys and censuses in the hope of targeting areas of improvement. A subgroup was assigned to focus specifically on editing and imputation procedures. This subgroup conducted an examination of these procedures in five separate economic programs, with the goal of identifying the size and timing of corrections as well as the sources of change (subject-matter analyst versus machine) to the edited data for each program (Shoemaker, 2006).

As a first step of this study, we developed a set of metrics applied to the reported data in comparison to the final tabulated data to assess the overall impact of the cumulative set of corrections to the reported data. Our final measures were a collaborative effort, developed by survey methodologists and subject-matter experts that refined and combined the proposed original measures as the study progressed. Our objective was to develop metrics that could be applied (with few modifications) to other similar programs.

In the next section, we give a brief background on the types of editing methods that are applied to economic data at the U.S. Census Bureau, providing general information on the processing system used and specific information on the StEPS editing and imputation **applications** for the program considered in this paper – the Annual Wholesale Trade Survey (AWTS). Following that, we provide a brief background on the Annual Wholesale Trade Survey. We next provide a discussion of the development of our proposed metrics, discussing our study limitations in a separate section. We present several illustrative results and conclude with some general comments and recommendations.

Oliver and Thompson (2007) describe the quality metric metrics presented here and apply them to data from two separate economic programs: the Annual Wholesale Trade Survey and the Annual Survey of Manufactures. This paper builds upon the evaluation presented in the earlier paper, refining these metrics by breaking out the earlier results to the industry level and adding an additional metric to evaluate the number of times the critical items in a record are edited/reviewed before attaining their final values.

## Editing of Economic Data at the Census Bureau

Although no universal definition of survey data editing exists, the following definition provided by the Federal Committee on Statistical Methodology's (FCSM) 1990 Working Paper 18 suffices:

*Procedures designed and used for detecting erroneous and/or questionable survey data (survey response data or identification type data) with the goal of correcting (manually and/or via electronic means) as much of the erroneous data (not necessarily all of the questioned data) as possible, usually prior to data imputation and summary procedures*.

---

[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 or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

Errors in the data can result from sources such as the respondent, the interviewer, and data capture instrument. The goal of **editing** is to detect these errors, not correct them. Items that fail edits via data editing rules are referred to an analyst for investigation and/or correction or are automatically changed (imputed) in an attempt to find values that do not fail the edit.

Editing can occur at the micro or macro level. Micro editing is generally the first stage of the edit review. Micro-edits check data at the record or questionnaire level. For most programs, records are automatically tested on a flow basis, with edit-failing values automatically designated for imputation (Thompson, 2007). Examples of micro-edits include:

- Range edits: edits that check if a reported value of a data item falls within a specified tolerance (i.e., $a \le x_i \le b$, where $x_i$ is the value of data item $x$ from reporting unit $i$, and $a$ and $b$ are predetermined limits).
- Ratio edits: edits that check if the ratio of the values of two data items fall within a specified tolerance (i.e., $a \le x_i/y_i \le b$, where $x_i$ is the value of data item $x$ from reporting unit $i$, $y_i$ is the value of data item $y$ from the same reporting unit $i$ and $a$ and $b$ are predetermined limits).
- Balance edits: edits that check if the sum of the details is equal to the reported total (i.e., $x_i + y_i = z_i$).

During the micro-review phase, analysts perform nonresponse follow-up and review the machine-imposed changes on key data items. Most programs allow analysts to override a machine impute with a value that is either derived from recontact with the survey unit, from a validated outside source, or from a simple (logical) correction. Moreover, most programs include selection criteria for this targeted analyst review or non-response follow-up such as reviewing processed data for all "large" cases or for all changed values from units with very large sampling weights.

Macro-edits test aggregate data to detect inconsistencies indicating possibly erroneous microdata. Checks may be done at various geographic levels, by strata, by industry, etc. A 1994 inventory of statistical practices in the economic area of the U.S. Census Bureau reported that macro-editing procedures varied widely across the economic area (King et al., 1994). Some surveys do a cursory review of data at aggregate levels, whereas others use automated methods. Many surveys review computed totals by comparing previous (published data) survey totals to current survey totals (Sigman, 2005).

A macro-edit failure requires research on the analyst's part to detect (and possibly validate) the source of the edit failure, again at the micro level. The micro-review of edit-failing records by subject-matter expert analysts often represents a high proportion of the total time spent on survey processing. As to what percent of the records that fail a micro level review are followed up on, this varies by survey. King et al. (1994) found that for the 85 economic surveys that they surveyed, "most of the percentages of unresolved cases lie between 10 and 20 percent."

The Census Bureau has developed two separate generalized processing systems for editing economic data: one for the Economic Census – called "Plain Vanilla"; the other for the surveys – called the Standard Economic Processing System (StEPS). See Oliver and Thompson (2007) for a comparison of both systems. In this paper, we only consider StEPS. StEPS contains modules for all survey processing activities, starting with data collection support (for example, printing labels), continuing through editing, data review and correction, imputation, and ending with calculation of estimates and variances and primary disclosure-avoidance processing. Currently, the Census Bureau uses StEPS to process approximately 80 economic surveys.

Here, we are studying the impact and efficiency of one survey's usage of the data editing and imputation modules in StEPS. StEPS has a one module for data editing and two modules for imputation. The usual order in which the modules are executed is first **simple imputation**, then **editing**, and finally **general imputation**.

The **StEPS simple imputation module** replaces missing or inconsistent input data values with values considered to be either reported data or of equivalent quality as reported data. An example of simple imputation is "data filling" wherein StEPS replaces a missing value for an item with a value derived from other data by subtraction. The difference between simple imputation and general imputation is that simple imputation uses logical edits whereas general imputation uses model-based imputation methodology.

The **StEPS editing module** performs automated detection of potential data errors. Subject-matter analysts must specify the "expected reported behavior" by entering edit definitions known as parameters. The StEPS editing module only identifies the edit failures – it does not change data. Executing edits interactively allows analysts to interactively make data corrections and then see if corrected data satisfies the edits (Sigman, 2001). There are several standard types of edit tests available in the StEPS editing module, such as required item tests or negative tests. However, the majority of the edits implemented by our case study survey are specified in survey rule tests. The survey rule tests, which form the backbone of the StEPS edit module do not have a predetermined structure and can combine more than one type of edit.

The **StEPS General imputation module** creates model-imputed values for items that have failed an imputation test, have been marked by an analyst to be imputed, are missing and require a value, or have been involved in a failed balance complex test. StEPS allows two very different types of imputation procedures: item imputation and imputation for balance complexes. Item imputation imputes or adjusts the value of a specific item that requires model imputation. Imputation for balance complexes imputes or adjusts the items in a balance complex so that the details will add to corresponding totals and/or subtotals.

**The Annual Wholesale Trade Survey (AWTS)**

The primary purpose of the Annual Wholesale Trade Survey **(**AWTS) is to provide detailed industry measures of sales and inventories for companies with employment that are primarily engaged in wholesale trade in the United States as defined by the North American Industry Classification System codes (NAICS). These include two main types of wholesalers:

- Merchant wholesalers that sell goods on their own. These include sales branches and offices (MSBOs) maintained by manufacturing, refining or mining enterprises for the purpose of marketing their products.

- Business-to-business electronic markets, agents and brokers (AGBRs) that arrange sales and purchases for others, generally for a commission or fee (first collected in the 2005 survey year).

Note: The studied 2003 AWTS data are based on the 1997 NAICS definitions. At the time of our analysis, the survey was called the Annual Trade Survey (ATS) – it is now called the Annual Wholesale Trade Survey (AWTS).

The AWTS is a mail-out/mail-back survey of about 8,000 wholesale businesses. The sample is drawn from the Business Register, which contains all Employer Identification Numbers (EINs)[2] and listed establishment [3] locations. Firms are first stratified by major kind of business and estimated sales. All firms with sales above applicable size cutoffs are included in the survey and report for all their wholesale industry EINs. In a second stage, unselected EINs are stratified by major kinds of business and sales, and randomly selected from each stratum (Source: U.S. Census Bureau website: http://www.census.gov/econ/www/wh0200.html; Wills, 2006).

For the AWTS study, we analyzed data pertaining to the following three critical items from the 2003 Annual Wholesale Trade Survey of Wholesale Distributors: Forms SA-42 and SA-42A: **Sales, Total Inventories,** and **Total Purchases**. Definitions for these critical items can be found at the above-referenced web site. Our analysis looked at Merchant Wholesalers excluding MSBOs.

The AWTS uses the StEPS edit module to identify respondent data that exhibit relationships that do not conform to expected behavior and to flag missing items for general imputation. The AWTS flags item for imputation or review based on two StEPS edit types: required data item and survey rule. The required data item edit will flag an item for imputation if the value is missing. A survey rule edit is a **free-form** edit test that validates inter-item relationships

[2] A number assigned to a business by the Internal Revenue Service for payroll reporting.
[3] The smallest business unit at which transactions take place or services are provided and payroll and employment records are maintained.

within an observation. Analysts manually resolve **all** edit-failures; complete unit non-response cases (i.e., "delinquents") are automatically machine imputed.

In addition to the machine edits, the AWTS analysts also frequently use independently-developed ad-hoc queries and write independent programs to identify records that may require extra review and correction. The AWTS program also uses the Hidiroglou-Berthelot edit (Hidiroglou and Bertholot, 1986) to detect outliers for selected current cell ratios based on the NAICS code and the type of unit (Merchant Wholesalers, MSBOs, and Agents & Brokers). The Hidiroglou-Berthelot edit takes the size and weight of the unit into account when evaluating whether its ratio is outlying.

**Edit Efficiency Metrics**

Working in conjunction with subject-matter analysts and other methodologists, we developed an initial set of metrics to assess the size of changes to the reported data as well as the source of change and type of change. In addition, we developed metrics to assess the effect of editing on the reported data. Before we provide these metrics, we first provide a set of definitions:

- **Item:** name of a field on a questionnaire.
- **Critical Item:** an item whose value is a key measure for a survey.
- **Reported Data:** an item whose value was reported directly by the respondent or derived indirectly from other reported items.
- **Final Edited Data**: the value of a data item used in the final tabulations.
- **Data flag**: an alphabetic flag that indicates the source of the data item (e.g., reported, corrected, imputed).

In the following sections, we define the metrics used to evaluate the efficiency of the editing:

**Metric 1:** The percentage of the records with reported values for a critical item whose value was changed by the edit/imputation program or by the analysts.

**Rationale:** Too many changes to the reported values for a particular item may indicate one of two problems with the edit rules: (1) Sometimes the edit rules are too restrictive, resulting in over-editing. As a result, data that are reasonable are changed to satisfy the edits; (2) Problems with the inquiry. For example, asking the respondent for data that are either not available or are kept in ways that make it difficult to supply information (Sudman et al., 2000). In either case, Metric 1 provides a means for further investigation and improvement of the edit process, long advocated by Deming and others (see Montgomery, 1991).

Definition:

- Let $x_{h1}, x_{h2}, x_{h3},\ldots, x_{hn}$ represent the $n$ reported, non-missing values of critical item, $h$.
- Let $x'_{h1}, x'_{h2}, x'_{h3},\ldots, x'_{hn}$ represent the corresponding $n$ final edited values of critical item $h$.
- Let $y_j = 1$ *if* $x_{hj} \neq x'_{hj}$
  $= 0$ otherwise

***Metric 1*** $= \dfrac{\sum_{j=1}^{n} y_j}{n} \times 100$ **= the percentage of the reported, non-missing values of critical item *h* that have been changed in the final edited version of the data.**

**Metric 2:** The percentage of changes to the reported values for a critical item (see Metric 1) that is attributable to analyst corrections versus machine corrections.

**Rationale:** Ideally, automated edits and imputation would detect and correct, respectively, most of the errors in survey returns; analysts would investigate and if possible, resolve the unresolved cases. Metric 2 provides information on the source of the final edited data.

**Procedure:** We consolidated the StEPS item data flags into the following general "source of change" categories:

- **Analyst Correction:** validated or verifiable corrections performed by the subject-matter analysts or clerks to fix apparent reporting errors.
- **Analyst Impute:** replacement value provided by a subject-matter analyst.
- **Machine Impute:** automated changes to the data via a variety of imputation programs.
- **No Change:** the reported value of a data item is equal to its final edited value.

We then assigned each record in the final edited version of the data for a critical item (see Metric 1 for definition) one of the *source of change* categories listed above. Records where the reported value for a critical item was equal to final edited value were assigned the last category. Records where the reported value was not equal to the final edited value were assigned one of the other *source of change* categories.

Definition:

- Let $x_{h1}, x_{h2}, x_{h3}, \ldots, x_{hn}$ represent the $n$ reported, non-missing values of critical item, $h$.
- Let $x'_{h1}, x'_{h2}, x'_{h3}, \ldots, x'_{hn}$ represent the corresponding $n$ final edited values of critical item $h$.
- Let $f_1, f_2, f_3, f_4$ represent the *source of change* categories that could be assigned to the final edited values (one category per record).
- Let $m_i$ represent the number of the $x'_{h1}, x'_{h2}, x'_{h3}, \ldots, x'_{hn}$ records assigned the $f_i$ source of change category. Hence, $n = m_1 + m_2 + m_3 + m_4$.

***Metric 2*** = $\frac{m_i}{n} \times 100 =$ **the percentage of the reported non-missing values of critical item h that have been changed due to source $f_i$.**

**Metric 3:** The relative size of change of the reported, non-missing values for a critical item in relation to the final edited values, broken out by *source of change* categories discussed in Metric 2. (See the discussion and Figure 1 below for an illustration).

**Rationale**: To understand the overall impact of the editing/imputation changes on the reported data, we need to examine the size of each change simultaneously, at both the macro and micro levels.

**Discussion**:
Figure 1 provides a partial listing of the output for Metric 3 for a given critical item for the AWTS. The complete analysis includes two additional *source of change* categories: Analyst Impute and Machine Impute. For brevity, we only listed two categories. For each *source of change* category, we divided the records into *size of change* size categories – a ratio comparison of the reported value to the final edited value. For example, **R/E** ≥ 900 means that for a given unique record, the reported value of that item is 900 or more times larger than the final, edited value. These category can be particularly useful for catching changes as "rounding errors" or "data slides" – records where the reported value was reported in the wrong units -- total dollars as opposed to thousands.

At the opposite end of the spectrum, the categories 1.0 < **R/E** < 1.1 and 1.0 < **E/R** < 1.1 are useful for determining the percentage of the records where the reported value is less than 10 percent larger than the final edited value (and vice-versa) – i.e., relatively small changes. If the percentage of records with these small changes is relatively large, it may indicate problems with the edit parameters or with the analyst review procedures. The category **R/E** = 1 indicates that the reported value of a data item is equal to its final edited value. It must be noted that these *size of change* categories are arbitrary and were designated by a team of survey methodologists and subject-matter experts.

Simply examining the ratio of changes to the reported data can be misleading: one must also examine the magnitude of the change, at both the micro and macro-levels. To illustrate the former, suppose there were $n$ records for a given critical item that were changed by analyst correction and fell into the 1.0 < **R/E** < 1.1 (as indicated on Figure 1). At first glance, these changes appear to be quite trivial. However, if a unit has a very large sampling weight, a small change in a dollar value could potentially have a large effect on the tabulated value.

To determine if this is the case, one can do the following:

1. Tabulate the weighted reported (column 4) and weighted edited values (column 5) of these *n* records, then determine the *percentage difference* between these values.
2. Find the absolute difference between the weighted reported and weighted edited value for each of the *n* records.
3. Sum up these absolute differences (column 6).
4. Find the average absolute difference by dividing the sum in column 6 by the number of records, *n*.

   Note: This average absolute value tells us the average shift in the weighted reported values in comparison to the final edited values for a specific change category. This type of average makes it easier to compare the impact of weighted changes across different change categories. Bear in mind though that it does not mean that all records within a given category had the same impact – it is just an average measure of change to the reported data over the *n* records in the category.

For the macro-level examination, we tabbed up the overall weighted reported values (totals row, column 4) and the overall weighted final edited values (totals row, column 5) for a given critical item and then found the percent between these sums (see totals row, column 6). This measure provides an overall indication as to the impact of changes to the reported data (a "bottom line"), but by itself does not indicate the source of the change or the size of the change. The former measures do. Additional statistics can be derived from Figure 1. For example, of the records whose reported values were changed (in comparison to the final edited values), what percent were changed due to Analyst Corrections versus Analyst Imputes versus Machine Imputes?

**Figure 1: Relative Size of Change to Reported Data and Source of Change for a given Critical Item**

| **Source of Change (1)** | **Size of Change (2)** | **No. of Records (3)** | **Tabulated Weighted Reported (4)** | **Tabulated Weighted Edited (5)** | **Percentage difference (6)** | **Sum of the Absolute Differences (7)** | **Average Absolute Difference (8)** |
|---|---|---|---|---|---|---|---|
| **Analyst Correction** | 1.0 < **R/E** < 1.1 | *n* | *x* | *y* | *(y – x)*100/x* | *z* | *z/n* |
| | 1.1 ≤ **R/E** < 9 | | | | | | |
| | 9 ≤ **R/E** < 90 | | | | | | |
| | 90 ≤ **R/E** < 900 | | | | | | |
| | **R/E** ≥ 900 | | | | | | |
| | 1.0 < **E/R** < 1.1 | | | | | | |
| | 1.1 ≤ **E/R** < 9 | | | | | | |
| | 9 ≤ **E/R** < 90 | | | | | | |
| | 90 ≤ **E/R** < 900 | | | | | | |
| | **E/R** ≥ 900 | | | | | | |
| INTENTIONAL BREAK | | | | | | | |
| **No Change** | **R/E** = 1 | | | | | | |
| **Totals** | | **Total 3** | **Total 4** | **Total 5** | **Perc. Diff.** | | |

**Metric 4:** The relative size of change of the reported, non-missing values for a critical item in relation to the final edited values, broken out on an industry (NAICS) level. See the discussion below and **Figure 2** below.

**Rationale**: Examining the effect of changes to the reported data in comparison to the final edited data across all NAICS is important (as discussed in Metric 3), as it provides an overall snapshot of the effect of editing/imputation on the reported data on the final survey tabulations. However, it is also important to examine the changes on an industry (NAICS) level, to assess whether trivial changes at the survey level are quite non-trivial at certain subcategory level, and whether data review and change patterns are consistent across all or most industries.

**Procedure:**
1. Separate the reported and final edited values by critical item and NAICS code.
2. Within each set, tabulate the weighted reported values and weighted edited values.
3. Calculate the ratio of the tabulated weighted reported and tabulated weighed edited values.
4. Categorize the ratio value into one of the *size of change* categories discussed in metric 3.

5. Place the results for a given critical item into a table such as Figure 2, shown below:

**<u>Discussion</u>**:
Column (1) indicates the name of the critical item; Column (2) provides a list of the various *size of change categories*, previously discussed in Metric 3; Columns (3) and (4) indicate the number and percent, respectively of NAICS industries whose ratio of the tabulated weighted reported values and tabulated edited values falls into a given *size of change* category; Columns (5) and (6) provide the weighted sum of all the reported and edited values that fall into a given *size of change category*; Column (7) provides the sum of the absolute differences between the weighted reported and weighted edited values and Column (8) provides what percent of the total of these absolute differences is accounted for by a given *size of change* category.

**Figure 2: Relative Size of Change of the Reported Data for a given Critical Item**

| **Name of Critical Item (1)** | **Size of Change (2)** | **No. of NAICS (3)** | **Percent of Total (3) (4)** | **Tabulated Weighted Reported (5)** | **Tabulated Weighted Edited (6)** | **Sum of the Absolute Differences (7)** | **Percent of Total (7) (8)** |
|---|---|---|---|---|---|---|---|
| | 1.0 < **R/E** < 1.1 | | | | | | |
| | 1.1 ≤ **R/E** < 9 | | | | | | |
| | 9 ≤ **R/E** < 90 | | | | | | |
| | 90 ≤ **R/E** < 900 | | | | | | |
| | **R/E** ≥ 900 | | | | | | |
| | 1.0 < **E/R** < 1.1 | | | | | | |
| | 1.1 ≤ **E/R** < 9 | | | | | | |
| | 9 ≤ **E/R** < 90 | | | | | | |
| | 90 ≤ **E/R** < 900 | | | | | | |
| | **E/R** ≥ 900 | | | | | | |
| INTENTIONAL BREAK | | | | | | | |
| | **R/E** = 1 | | | | | | |
| | | **Total (3)** | **100.0%** | **Total (5)** | **Total (6)** | **Total (7)** | **100.0%** |

**<u>Metric 5</u>:** How many times are the records for each critical item changed by analysts and/or machine imputes before attaining their final value (i.e., how many "cycles" or passes of editing/imputation does each key item undergo)?

**<u>Rationale</u>:** The reported value of a critical item may be changed several times by analysts and/or machine imputation before attaining its final value. By determining the numbers of edit passes it takes to arrive at the tabulation values, we have a baseline measure for future improvement of the process.

**Study Limitations**

We analyzed changes to the "reported" data in comparison to the final edited data. For the AWTS data, the reported data may not have been **originally** provided by the respondent (e.g., it may have been replaced after analyst contact). Additionally, we only examine data from the 2003 data collection, and the results presented here may not be representative of results that would have been obtained with other survey years' data.

**Results**

**Illustration of Metric 1**
Table 1 illustrates Metric 1 and provides a comparison of the tabulated weighted reported values and the tabulated final edited values for the critical items specified. We demonstrate the usage of this table by examining the results for **Sales.** For **Sales**, there were 4,819 records that had reported values (as identified by the data flag). Of these 4,819 records, 238 (4.9 percent) were changed as a result of editing and imputation. Correction of these relatively few records resulted in a 93.9 percent reduction of the overall reported amount. The percent reductions in overall reported amounts for **Purchases** and **Inventory** were 56.5 percent and 98.8 percent, respectively.

**Table 1: Illustration of Metric 1**

| Critical Item | No. of records with reported values | No. of records changed | Percent | Reported Amount (Weighted) | Edited Amount (Weighted) | Percent Difference |
|---|---|---|---|---|---|---|
| Sales | 4,819 | 238 | 4.9% | $41,156,147,469,827 | $2,156,439,761,795 | – 93.9 % |
| Purchases | 4,628 | 403 | 8.7% | $4,486,157,565,225 | $1,953,140,854,409 | – 56.5% |
| Inventory | 4,334 | 326 | 7.5% | $21,392,659,055,594 | $256,920,242,139 | – 98.8% |

**Illustration of Metric 2**

Table 2 illustrates Metric 2 and provides information on the magnitude of change in the reported data. Notice that for **Sales**, a total of 238 records (221 + 15 + 2) were changed as a result of analyst corrections, analyst imputes, or machine imputes. As expected, analyst corrections accounted for the overwhelming majority of the final changes (92.9 percent). This follows from the AWTS data processing procedures: recall that analysts manually review and correct the majority of edit failures, and that only reported data items are considered in these metrics (imputed data items for non-respondent cases are excluded). Imputation, performed by analysts and machine accounted for 6.3 and 0.8 percent, respectively. On average, the reported dollar amount of the 221 records corrected by the analysts (for apparent reporting errors – mostly rounding errors) was shifted by $175,545,287,020 (in terms of absolute value).

The average shift in the reported amount for these 221 records corrected by analysts was much higher than the average shift for the 15 records that were imputed by analysts by a ratio of 269 to 1. Two records had their final changes imputed by machine. Their average shift in dollars was $55,733,569. An inspection of the other two critical items, **Purchases** and **Inventory** reveal similar results. Analysts made the majority of the final changes to these records also. An inspection of the corresponding data flags revealed that records with "rounding errors" accounted for a large average shift in the overall reported amount for each critical item.

**Table 2: Illustration of Metric 2**

| Critical Item | Source of Change | No. Records Changed | Percent of Total | Average Absolute Difference | Ratio of AC to AI and AC to MI |
|---|---|---|---|---|---|
| **Sales** | Analyst Corrections (AC) | 221 | 92.9% | $175,545,287,020 | ----- |
| | Analyst Impute (AI) | 15 | 6.3% | $653,755,569 | 269/1 |
| | Machine Impute (MI) | 2 | 0.8% | $55,733,560 | 3150/1 |
| **Purchases** | Analyst Corrections (AC) | 363 | 90.1% | $7,404,687,898 | ----- |
| | Analyst Impute (AI) | 37 | 9.2% | $289,041,007 | 26/1 |
| | Machine Impute (MI) | 3 | 0.7% | $78,670,750 | 94/1 |
| **Inventory** | Analyst Corrections (AC) | 285 | 87.4% | $74,196,357,889 | ----- |
| | Analyst Impute (AI) | 15 | 4.6% | $73,367,584 | 1011/1 |
| | Machine Impute (MI) | 26 | 8.0% | $38,766,412 | 1914/1 |

**Illustration of Metric 3**

Recall that Metric 3, as illustrated in Figure 1, provides information about the sources of change to the reported data in relation to the final edited values as well as the impact of these changes at both the micro- and macro-levels. Because these tabulations are quite large, we do not present the complete findings for the AWTS in this paper (they are available upon request from the authors). Instead, we highlight the major findings obtained from this metric.

As shown in Table 2, a very small percentage of the reported values for **Sales** were changed (only 238 of the 4,819 records, i.e., 4.9 percent) were changed. Of these 238 records, two records (0.8 percent) were changed by machine imputes. Analyst Corrections accounted for the majority of the changes (221 records or 92.9 percent). Analyst Imputes accounted for 6.3 percent of the changes. Most of the analyst changes fell into the $1.1 \leq$ **R/E** $< 9$ and $1.1 \leq$ **E/R** $< 9$ categories. Relatively few fell into the **R/E** $\geq 900$ and **E/R** $\geq 900$ categories, but changes to these records had the greatest impact on the tabulated data. These analyst changes were mostly the correction of rounding or keying errors (e.g., dropped digits), which can have high impact on tabulated weighted values but can be easily identified and corrected by an experienced analyst. Quite reasonably, the changes to records in the **R/E** $\geq$ 900-change category had the greatest impact on the tabulated data. These corrections were mostly carried out to correct rounding (divide by 1000) errors. The analysts did not work on too many cases (six cases in total) in the 1.0 < **R/E** < 1.1 or 1.0 < **E/R** < 1.1 change categories. The results for **Purchases** and **Inventory** were similar.

**Illustration of Metric 4**
In the previous sections, we discussed the impact of changes to the reported data for **Sales**, **Purchases**, and **Inventory** on the final **survey** tabulations. In the following sections, we examine these changes on an industry (NAICS) level. Again, we will not present the results in the tabular form illustrated, as there will be too many tables. Instead we will present the highlights of our findings for each of the critical items, and then make some generalizations across items.

**Sales.** With respect to the ratio of the tabulated weighted reported values to the tabulated weighted edited values, we found that 51.5 percent of the NAICS industries had ratio values that fell into the 1.1 ≤ **R/E** < 9 and 1.1 ≤ **E/R** < 9 *size of change* categories. To investigate the relative contribution of these changes on the overall tabulation, we first removed the following *size of change* categories: **9 to 90**, **90 to 900**, and **900 or more,** whose records exerted a large influence on the tabulations and whose categories tend to be indicative of "rounding" errors. With these categories removed, we found that the percentage difference between the tabulated weighted reported and tabulated weighted edited values was –27.3%, representing a reduction in the overall weighted reported values. This indicates that the cumulative impact on the final survey tabulation of the cases whose ratio fell into the 1.1 ≤ **R/E** < 9 and 1.1 ≤ **E/R** < 9 *size of change* categories was not trivial.

We also examined those industries where the overall weighted reported values to the overall weighted edited values fell into the 1 ≤ **R/E** < 1.1 and 1 ≤ **E/R** < 1.1 categories. Approximately 5.9 percent industries fell into these categories. The percentage difference between the total weighted reported and total weighted edited values for these industries was relatively small, 4.2 percent. This demonstrates that the analysts are **not** making "small" (fine-tuning) changes to the data in order to satisfy the **Sales** edits, providing some evidence for the efficiency of the employed edits for **Sales.**

**Purchases**. With respect to the ratio of the tabulated weighted reported values to the tabulated weighted edited values, we found that 64.7 percent of the NAICS industries had ratio values that fell in the 1.1 ≤ **R/E** < 9 and 1.1 ≤ **E/R** < 9 *size of change* categories. After removing the **9 to 90**, **90 to 900**, and **900 or more** *size of change* categories, we found that the percentage difference between the tabulated weighted reported and tabulated weighted edited values was –22.8%, representing a reduction in the overall weighted reported values. Again**,** the overall impact of these changes is not trivial at the survey level. Moreover, this *size of change* pattern appears to be fairly consistent across industries.

Approximately 7.4 percent of the NAICS industries had ratios that fell in the 1 ≤ **R/E** < 1.1 and 1 ≤ **E/R** < 1.1 *size of change* categories. The percentage difference between the total weighted reported and total weighted edited values for these industries was relatively small, 1.9 percent. This demonstrates that the analysts are **not** making "small" (fine-tuning) changes to the data in order to satisfy the **Purchases** edits, providing some evidence for the efficiency of the employed edits for **Purchases.**

**Inventory.** With respect to the ratio of the tabulated weighted reported values to the tabulated weighted edited values, we found that 55.9 percent of the NAICS industries had ratio values that fell in the 1.1 ≤ **R/E** < 9 and 1.1 ≤ **E/R** < 9 *size of change* categories. With the **9 to 90**, **90 to 900**, and **900 or more** *size of change* categories removed, we found that the percentage difference between the tabulated weighted reported and tabulated weighted edited values was –36.6%, representing a reduction in the overall weighted reported values. As with **Sales** and **Purchases,** the overall impact of these changes is not trivial at the survey level. Moreover, this *size of change* pattern appears to be fairly consistent across industries.

Approximately 10.3 percent of the NAICS industries had ratios that fell in the 1 ≤ **R/E** < 1.1 and 1 ≤ **E/R** < 1.1 *size of change categories*. The percentage difference between the total weighted reported and total weighted edited values for these industries was relatively small, 3.0 percent. This demonstrates that the analysts are **not** making "small" (fine-tuning) changes to the data in order to satisfy the **Inventory** edits, providing some evidence for the efficiency of the employed edits for **Inventory.**

**Illustration of Metric 5**
Our final metric examines the number of times the records for each of the three critical items (**Sales**, **Purchases**, and **Inventory**) were subjected to editing and imputation. Our goal was to determine the impact of multiple edit passes on the tabulations. How many cycles of editing and imputation do the records for each critical item undergo before attaining their final values?

Interpreting our results was somewhat confounded by the AWTS imputation procedures. This survey "re-imputes" each imputed case as the imputation base changes. This occurs throughout the survey processing cycle, as new cases are received and as edit-failing cases are reviewed and changed via (verified) analyst correction. Excluding "delinquent" cases from the computations helped us interpret the results.

Aside from machine-imputed items, the majority of substantive changes are made within three passes for the three critical items. During the first edit cycle, analyst corrections of the records -- particularly those with rounding errors -- resulted in a dramatic decrease in the overall tabulated amount. The number of changed records decreases steadily with each editing cycle with very few records being edited after the third cycle. The large average change (in the tabulations) per record in the first edit cycle dropped substantially in subsequent edit cycles and essentially leveled off beginning with the sixth edit pass for **Sales** and **Purchases** and the eighth edit cycle for **Inventory**. These results are very consistent with those presented from Metrics 3 and 4, namely that the analysts are first correcting large and obvious errors, then using subsequent reviews to search for "missed" errors that were previously masked. Because the number of edit failing records is rather small, it is possible for the AWTS analysts to thoroughly review each case. Unfortunately, it is impossible to recreate the existing survey data at any given cycle since they are continuously updated, so we cannot directly assess the impact on the tabulations of these multiple cycles. However, the evidence does seem to indicate that if time were of the essence, comparable quality could be obtained with fewer reviews.

**Conclusion**

In this paper, we present a set of tabulations and metrics designed to assess the combined impact of editing, imputation, and analyst review procedures on economic data. We illustrated the usage of these measures with the AWTS, examining them at both the survey level and at the individual industry level. We also examined the number of edit passes used to correct each reported item in an attempt to assess the "value-added" from repeatedly subjecting reported data to editing and analyst review.

What we did **not** find is as important as what we did find. The majority of our findings were not unexpected: for this survey, analyst correction accounted for the majority of changed to reported data (a consequence of the survey processing procedures) and the correction of rounding errors accounted for the most substantial change in the tabulated values. Only a handful of the cases fell into this category, and a large proportion of such cases were resolved via human interaction. Rounding errors are often easily identified by comparison to non-dollar items or to prior period tabulated data values and could be easily corrected by machine.

After removing these rounding cases from our metric calculation, we found that the majority of the changes to the records fell into the $1.1 \leq \mathbf{R/E} < 9$ and $1.1 \leq \mathbf{E/R} < 9$ *size of change* categories, regardless of data item – meaning that the ratio of the reported value in comparison to the final edited value (and vice-versa) was between 1.1 and 9. Changes to these records across all NAICS and within NAICS had an important impact on the final tabulations. On the opposite end of the spectrum, changes to records that fell into the $1 \leq \mathbf{R/E} < 1.1$ and $1 \leq \mathbf{E/R} < 1.1$ categories were relatively few and their impact on the final tabulations was much less – across all NAICS and within NAICS. This pattern was fairly consistent across industries – in other words, we **did not find** that analyst changes were focused on a handful of industries, or that analysts were systematically making a high proportion of small changes to any data item. This indicates that the AWTS edits are effectively isolating outlying values in their industries.

Granquist (1995) proposes examining the number edit cycles (passes) employed for each data item to assess the overall efficiency and value added of the analytical review phase. We did this with the AWTS data and found that most cases were completely resolved within three cycles. The analysts handled the edit failures and machine handled the delinquent cases that had to be imputed. Again, we **did not find** any evidence of "wasted effort" in this cycling, since it appears that the analyst procedures begin with the large and obvious errors and funnel down to the not-insubstantial but less detectable reporting errors that were previously masked.

We believe that the process presented here is as important as the survey results. We have developed a standard set of metrics that can easily be applied to other programs and that can easily be interpreted. A principle of statistical quality control is to repeatedly apply the same metrics to data at various phases to monitor an ongoing process. Our presented metrics can be applied to survey data throughout the survey processing cycle at given time intervals and can be used to monitor quality and to evaluate "value added" from repeated processing cycles on a flow basis. Although the utility of such a review for AWTS (which has a very small number of edit failures) is debatable, it could prove extremely useful for a survey that has more edit failures, an inconsistent edit failure pattern, or several distinct survey processing phases.

**Acknowledgements**

The authors thank Kari Clark and Rita Petroni for their useful comments on an earlier version of this paper.

**References**

Federal Committee on Statistical Methodology (1990). Statistical Working Paper 18: Data Editing in Federal Agencies.

Granquist, Leopold (1995). Improving the Traditional Editing Process. Business Survey Methods. New York: John Wiley and Sons, pp. 385-481.

Hidiroglou, M.A. and Bertholot, J.M. (1986). Statistical Editing and Imputation for Periodic Business Surveys. Survey Methodology, **12**, 73-83.

King, C. and Kornbau, M. (1994). Inventory of Economic Area Statistical Practices: Phase 2: Editing, Imputation, Estimation, and Variance Estimation. ESMD Report Series ESMD – 9401. *Internal U.S. Census Bureau Document, available from the Office of Statistical Methods and Research for Economic Programs*.

Montgomery, D.C. (1991). Introduction to Statistical Quality Control. New York: John Wiley and Sons.

Oliver, B.E. and Thompson, K.J. (forthcoming in 2007). Quality Metrics for Assessing the Impact of Editing and Imputation on Economic Data. *Proceedings of the Third International Conference on Establishment Surveys*, American Statistical Association.

Shoemaker, D. (2006). Business Process Improvement Team: Edit Efficiency Team Evaluation Studies. *Internal U.S. Census Bureau Document, available from the Economic Program Coordination Division*.

Sigman, R.S. (2005). Statistical Methods Used to Detect Cell-Level and Respondent-Level Outliers in the 2002 Economic Census of the Services Sector. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Sigman, R. (2001). Editing and Imputation in a Standard Economic Processing System. *Proceedings of Statistics Canada Symposium 2001: Achieving Quality in a Statistical Agency: A Methodological Perspective.*

Sudman, S., Willimack, D., Nichols, E., and Mesenbourg, T. (2000). Exploratory Research at the U.S. Census Bureau on the Survey Response Process in Large Companies. *Proceedings of the Second International Conference on Establishment Survey (Survey Methods for Businesses, Farms, and Institutions).*

Thompson, K.J. (forthcoming in 2007). Data Editing, a Little Imputation, and Coding. *Proceedings of the Third International Conference on Establishment Surveys*, American Statistical Association.

Wills, K. (2006). The Annual Wholesale Trade Survey: Survey Characteristics Relevant to Editing. *Internal U.S. Census Bureau Document, available from the Services Sectors Statistics Division*.